PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) **International Patent Classification** 6 : **H04N 7/14** — **A1**

(11) **International Publication Number:** **WO 99/57900**

(43) **International Publication Date:** 11 November 1999 (11.11.99)

(21) **International Application Number:** PCT/US99/09515

(22) **International Filing Date:** 1 May 1999 (01.05.99)

(30) **Priority Data:**
60/084,001 3 May 1998 (03.05.98) US

(71)(72) **Applicant and Inventor:** MYERS, John, Karl [US/US]; 300 Legacy Drive #1037, Plano, TX 75023 (US).

(74) **Agents:** CANNIFF, Brian, P. et al.; Elman & Associates, 20 West Third Street, P.O. Box 1969, Media, PA 19063–8969 (US).

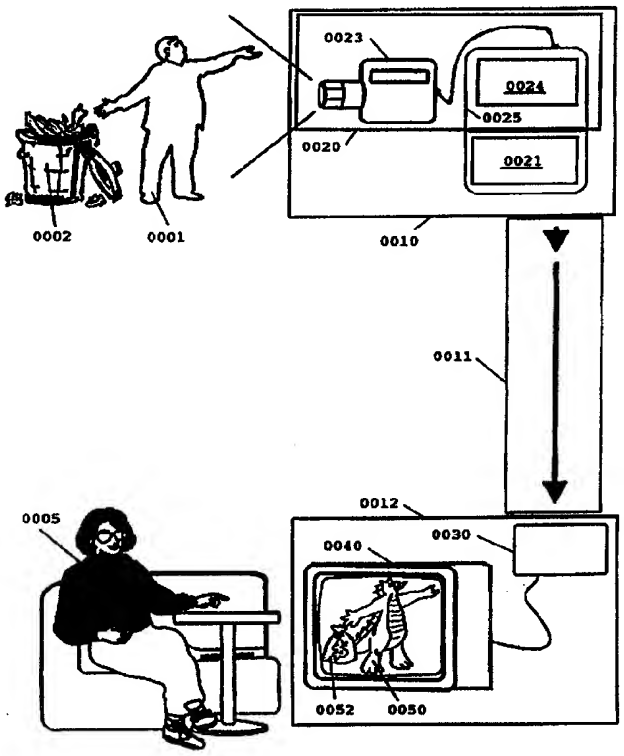
(81) **Designated States:** AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) **Title:** VIDEOPHONE WITH ENHANCED USER DEFINED IMAGING SYSTEM

(57) **Abstract**

An improved videophone comprising an information transfer device permitting one or more users to communicate to one or more viewers while allowing the appearance, sound, motion, and characteristics of the individual users and the users' environments to be changed, replaced, augmented, modified, filtered, enhanced, or deleted. The improved videophone comprises an imaging system that acquires sensory information from the one or more users and/or the users' environments, and represents at least the essence of this information; a distribution channel that relocates the essential information from the users' locale to the viewers' locale; a presentation system that takes the essential information and creates for the one or more viewers a presentation that can change, replace, augment, modify, filter, enhance, and/or delete features, components, portions, and/or parts or wholes of the sensory appearances of the one or more users and the users' environments. The resulting one–way communications circuit can be used to construct two–way communications circuits, multi–user consensual cyberspaces, an e–mail system comprising improved videophone e–mail sending and e–mail receiving stations, an improved videophone answering machine, improved videophone TV broadcasting and receiving stations and other embodiments. The invention can be embodied using various imaging, distribution, and presentation technologies, in various connection topologies. The result can be used for a multitude of purposes, including such things as communication, entertainment, education and instruction, call center functions and technical support, sports, news, drama, art, or play.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

VIDEOPHONE WITH ENHANCED USER DEFINED IMAGING SYSTEM

TECHNICAL FIELD

This invention relates to videophones, specifically to an improved computer-assisted videophone that is used to communicate between ordinary people in ordinary situations without using intrusive hardware, and that allows users to change, enhance, or replace the visual and auditory images of themselves and their environments in a fantastic manner.

BACKGROUND

Failure of the market for current art videophones Videophones have had a long history. AT&T first built a Picturephone™ in its lab in 1956. This made its public debut at the 1964 World's Fair, and was introduced as a mass-market product in 1970, but it never caught on. Around 1990 AT&T came out with a newer consumer model which used ordinary phone lines to transmit both video and sound. The video picture was the size of four postage stamps. This too was not a commercial success. Not to be stopped, in 1993 AT&T came out with the Model 70 videophone that was integrated with a desktop computer, which again failed. Apple introduced its computer VideoPhone Kit in 1996, and the same year Microsoft introduced its software NetMeeting product. Even though commercial videophones have been available for almost 30 years, they have as of yet failed to be adopted for mass usage by the general population.

Current art presents unenhanced true appearance of users' faces, environments, and voices. Ordinary consumers agree that there are three main problems inherent in using a videophone:

1) The videophone user (being projected) does not want the person at the other end of the line (the "viewer") to see the "background" environment surrounding the user, as it is probably messy. This is especially a problem with unexpected incoming video calls. Anything less than a perfect environment will cause the user to lose face and feel embarrassed.

2) The user does not want the viewer to be able to see the face, hairstyle, and clothing of the user unless the user is impeccably made-up, groomed, and dressed. This is fine for professional fashion-models and newscasters; however, the ordinary housewife won't answer the phone if her face is unenhanced.

3) Some users don't want their true face to be seen at all, or want their true voice to be heard at all. However, answering the phone with a blank screen or with silence is also rude.

The combination of these seemingly insurmountable problems has prevented videophones from catching on in the general marketplace. Curiously, although these problems have been known by videophone marketing people for decades, they have been unrecognized and unexplored by engineers. A usable videophone has been a long-time dream of the population, and yet the large majority of ordinary people refuse to use one because these problems are viewed as inevitable.

Current art does not have the low bandwidth required for 640x480 "full-screen" 30 frames-per-second video Current software-based videophones that use regular telephone lines for communication present small, jerky pictures because they cannot compress the 73,728 Kbits/second required to simply transfer the video portion of the signal into the 28.8 Kbits/second modems prevalent in the installed base of the consumer population. There simply is not enough bandwidth when the entire picture and voice are being transmitted literally, even with excellent compression.

This problem also applies to devices not used as two-way videophones but simply used as one-way "TV"-style movie-viewing devices. No cell-phone TVs or wrist-watch TVs exist that get their signal simply over the telephone network.

This same problem has prevented movies of people, and of environments, from being widely used on the World Wide Web. The bandwidth required to transmit full-screen full-speed streaming movies over common phone lines simply is not there with current technology.

Bandwidth is also a problem when it comes to the size of a file required to be forwarded and stored for video or voice e-mail. Because the current bits-per-second for video mail is so large, the resulting mail files are huge and cannot be stored and archived easily. Mail files that are too large can also clog servers.

Current videophone art is restricted to actual sound and video images, and thus cannot compensate for misplaced microphones or cameras, or cameras with bad lens angles Tiny, up-close devices such as videophone watches, videophone cell-phones, and videophone palmtop computers, etc., by necessity acquire video and sound images from unnatural positions that are too close to the user and at bad angles. Even regular videophones using large screens must currently place the camera above the screen, requiring the user to choose between making eye contact with the outgoing camera or the incoming screen image of the remote user, but not both. Tiny up-close devices also require the use of large-angle ("fisheye") lenses in order to image the entire face of the user. The results from these problems are ugly.

Current avatars in cyberspaces are controlled by unnatural and unintuitive buttons and commands A number of 2D and a few 3D cyberspace environments already exist. Avatars, consisting of animated 2D or 3D bodies and faces that represent users in these environments, are currently controlled by keyboard commands or by mousing dials or buttons on the screen. The resulting communications are slow, difficult for the naïve user, and do not reflect the actual bodily poses and facial expressions of the user. Sound is again transferred literally in an undisguised and unenhanced manner.

Current technologies for motion capture are bulky, inconvenient, technical, and extremely expensive, and cannot be used by ordinary people.

Current games cannot be used as general-purpose players for movies composed from motions and expressions captured from actors The few software games that incorporate motion-capture files use them in a hard-wired manner. The viewer is not allowed to swap out the script & display information for another set, and have a completely different game work. There are no players that allow linear TV-style movie shows or nonlinear branching interactive adventures that have been derived from actors' motions and expressions to be displayed in a general-purpose manner. No system reads in a script, a set of files, and a set of essential information taken from actors to enable reconstituting the movie or game of the week.

Current technologies for environment replacement are inconvenient, technical, and extremely expensive; they do not work with essential information over a distribution channel, and thus cannot be modified by the viewer or remodified by the user; and they cannot be used by ordinary people

Previous environment-deletion and modification systems have used a blank blue-screen or green-screen as the environment to be deleted. This includes systems for television weather forecasters and movie actors. Both blue-screen and green-screen environments have several drawbacks, including: a very expensive custom studio carefully and completely painted with exactly the correct shade of optical blue must be used; for this reason common homes cannot be used; a blue or green tinge reflects off the walls and colors the face of the user unnaturally; the user must pay attention to their clothing and must not wear accessories that have blue or green in them, for fear of looking invisible at that point; expensive professional lighting should be used; and typically special custom hardware is required to delete the environment. The systems cannot work with a typical, everyday, unprepared home. However, the present art's main problem is that the systems perform compositing at the locale of the user, and then ship a finished product over the distribution channel. This is true in the case of television weather-person broadcasting shows. This is also true in the case of movies being distributed to movie theaters. The viewer cannot choose to change the enhancement to replace the environment with something the viewer chooses; and the user cannot change his or her mind and replace the environment with something else. In addition, enhancements are limited to replacement, and do not consider such things as augmentation, etc. The general population does not have a room painted completely blue and cannot use this technology for everyday videophone calls.

Recently, a number of smaller companies have come out with first-generation Internet videophones. A camera captures a series of images of the user. The images are digitized into a computer, compressed, sent across a communications link to a viewer's computer, uncompressed, and displayed for the viewer to watch. A similar simultaneous process handles the sound. An important point is that the entire image as it exists is shipped; no attempt is made to separate the user from the visual background. As a result, much data is required in this communications process. Even with very good image compression techniques, when using 28.8 Kb/sec modems to connect over the Internet only a few frames can be transmitted per second, and the resulting displays are still 1/4 or 1/2 of a standard 640x480 screen in both width and height. Because they compress the entire image, such approaches are intrinsically limited in bandwidth. These devices do not solve the problem of how to augment, enhance, or replace the user's or the background's appearance.

A very small population of basic researchers in the field of machine vision have started to examine the problem of tracking human body or human facial motion for its own sake. Like visually tracking a satellite, or tracking parts on a conveyor belt, this research has been done mostly as a theoretical exercise, with the hope that it might be used to further communication between a person and a computer.

Brenda Laurel built a virtual-reality research workstation for concept exploration and performance-art purposes. Actors inside special body suits and wearing 3D display goggles moved around. The full-body suits were instrumented with joint sensors. A cumbersome cable connected the actors to a single computer. The actors each controlled a virtual animal, which did not look like them, inside a 3D virtual cyberspace world. No attempt was made to augment or enhance existing appearances. The system generated a video-tape of the actors' performance in the virtual playground as its output; although the blinded actors were kept separate by necessity to keep them from crashing in to each other, no attempts at building a system for remote communication were performed. The input system was definitely intrusive, requiring very expensive laboratory equipment and much time to suit up. Expert technicians were required to run and maintain the equipment.

Only until recent years have the technology for performing visual face-tracking and body-tracking, and the PC color cameras necessary to do vision research, become widely available to the general public. Existing virtual-actor systems use a cumbersome imaging system to change the presented appearance of a user, but do it all on one machine and have no distribution channel.

The graphics technology required to build a 3D virtual environment or a 3D model of a face has been around since at least 1979 (see Newman and Sproull [15]), and videophones have been around since at least 1964. Even Microsoft's NetMeeting was introduced in 1996. During all this time, no one else has thought to address the fundamental problem preventing videophones from catching on in the mainstream market—the fact that it is undesirable for a viewer to see an actual presentation of the user's environment and the user's face in an unenhanced manner.

A number of vision researchers have worked on abstracting the image of a user and re-presenting it. However, they concentrate exclusively on reconstructing the image of the actual user, and not performing changes such as replacing it, augmenting it, or deleting parts of it. Virtual reality companies have constructed a system consisting of a distribution channel with a presentation system. However, they have no vision imaging system for input, relying instead on clumsy buttons.

Each of the references cited anywhere in this document is hereby incorporated herein in its entirety. However, to the extent that the explicit or implicit definitions of terms used in any such reference are inconsistent with any implicit or explicit terms herein, the term as defined herein shall prevail. The characterization of these documents is not intended as an admission that the document discloses more or less than it actually does in fact disclose. These documents, taken as a whole, help demonstrate how the present invention improves upon that which is already in the prior art and uses some prior art in inventive ways.

[1] Videoconferencing & Interactive Multimedia: The Whole Picture, by Trowt-Bayard and Wilcox, Flatiron Publishing, 1997. This book covers the history and current state-of-the-art of videophones. It also discusses applications, various distribution channels, presentation devices, and compression. The compression algorithms, systems, and products discussed cannot change the presentation of the user to something new. The book provides very good information on how to deploy a constructed system.

[2] Facial Animation: Past, Present and Future, by Terzopoulos et al., Siggraph '97, pp. 434-436. This short paper provides an overview of the state of the art in facial animation. It discusses different techniques for creating presentations of faces from actuation variables, and shows that building faces that move based on essential information is well within the state of the art. However, none of the many examples discussed in the paper use a distribution channel to carry essential information to a presentation system; in each system, all of the graphics are designed and created on the same machine. Similarly, none of the systems attempts to acquire images of a user for transmission and presentation.

[3] View Morphing, by Seitz and Dyer, Siggraph '96, pp. 21-30. This paper describes in detail the mathematics and algorithms for perspective view warping of a single face, and subsequent perspective view morphing between two eigenfaces. The system only works with two eigenfaces at a time. The system only works with the actual images of users, or the actual images of two users; no attempt is made to use the parameters of one image to construct a replacement image from a different set of eigenfaces. The system only works on one computer; no distribution channel is used for communication. The system does not do recognition of eigenvalues of an acquired image from previously-determined eigenfaces.

[4] Simulating Humans, by Badler et al, Oxford University Press, 1993. A comprehensive book on how to build computer-graphic presentations of jointed human bodies, and drive them using actuation variables. The bodies are controlled by an automatic computer program. No attempt is made to

control the bodies from image acquisition of actual users, nor is a distribution channel used to separate a user from a viewer.

[5] PC Telephony--The complete guide to designing, building, and programming systems using Dialogic(tm) and related hardware, 4th edition, by Edgar, Flatiron Publishing, 1997. A technical book that covers the actual implementation of computer telephony distribution channels.

[6] Automatic Construction of 3D Human Face Models Based on 2D Images, by Tang and Huang, 0-7803-3258-X/96 IEEE conference, 1996. This paper goes over the method of how to create a 3D model of a user's face using input images, and then map textures from photographs of the user's face onto the resulting image. 32 feature points are used. Although the work is performed on a single computer, the paper mentions the possibility of using this model in a videophone. However, no attempt nor mention is made of changing the appearance of the user nor of the environment in the resulting presentation.

[7] Extraction of Facial Sketch Image Based on Morphological Processing, by Li and Kobatake, 0-8186-8183-7/97 IEEE conference, 1997, pp. 316-319. This paper shows how to extract the computer-vision features of a human face from visual input. 38 feature points and 18 characteristic curves are extracted for each face, giving as output a representation of the actual user. No attempt is made to change the user's representation into something different, nor to communicate the representation over a distribution channel.

[8] Light Field Rendering, Levoy and Hanrahan, Siggraph '96. This paper discusses the four-dimensional space representation of images of objects, how it can be acquired and rendered. No attempt is made to change the image, to acquire pictures of users, nor to use a distribution channel.

[9] A principal component based probabilistic DBNN for face recognition, Shen, Fu, et al, 6-7803-3258-X/96 IEEE conference 1996. This paper uses a decision-based neural network and an eigenface system to recognize faces of users. No attempt to transmit the resulting eigenvalues is performed.

[10] Motion Estimation of Lips in pronouncing Korean Vowels based on Fuzzy Constraint Line Clustering, by Jung and Kim, 0-7803-3258-X/96 IEEE 1996. Demonstrates dynamic tracking of human mouths.

[11] Description of Eye Figure with Small Parameters, by Mukaigawa and Ohta, 07803-3258-X/96 IEEE 1996. Takes a picture of a user's eye, abstracts essential information describing the eye, and reconstructs a photograph of the same eye from only the essential information. No attempt is made to change the presentation. No attempt is made to communicate the information over a distribution channel.

[12] Face Localization and Facial Feature Extraction based on Shape and Color Information, by Sobottka and Pitas, 5-7803-3258-X/96 IEEE 1996. Given no previous windowing, this paper recognizes and

extracts the face of a user from anywhere in an input image, using color and local shape information. Then the eyes and mouth are located. No attempt is made to distribute this information.

[13] How to Program a Virtual Community, by Powers, Ziff-Davis 1997. Goes over the applied technology of constructing and running a cyberspace. Discusses avatars and virtual costumes for people, also distribution channels, topologies, and third-party suppliers. No mention is made of acquiring images from the actual user nor an imaging system, nor using resulting information to drive the avatars or costumes.

[14] Networks and Telecommunications: Design and Operation, by Clark, John Wiley and Sons, 1991. Goes over construction of a telecommunications network, including videophones. Does not mention changing the appearance of the user.

[15] Principles of Interactive Computer Graphics, by Newman and Sproull, McGraw Hill, second edition 1979 (first published 1973). One of the first comprehensive books on computer graphics. Includes pictures of computer-generated environments such as an airport landscape or a full kitchen table, along with pictures of a detailed shaded-polygonal human face model (p.397). Demonstrates mathematics for computer-graphic augmentation of videophone images. Does not discuss methods for distribution of essential information from one computer to another.

[16] Official Microsoft NetMeeting Book, by Bob Summers, Microsoft Press, 1998. An excellent introduction to the technology and usage of PC-based videophones. Presents the Microsoft NetMeeting™ videophone product in depth. NetMeeting uses PCs on the sending and receiving sides to transmit audio-visual signals over TCP/IP channels, including Direct Broadcast Satellites (DBS), ISDN or T1 lines, wireless or cable modems, Ethernet intranets and the Internet. Point-to-point and star topologies are also discussed. This demonstrates that distribution channels for videophones are well within the current state of the art. NetMeeting sends the entire (compressed) picture in a live fashion. It does not allow changing the appearance of the users nor the users' environments.

[17] Building Internet Applications with Visual C++, by Gregory, et al, Que Corporation, 1995. Describes how to build e-mail applications and programs that use sockets and data streams under Windows, from the ground up.

[18] Advanced Programming in the Unix Environment, by Stevens, Addison-Wesley Professional Computing Series, 1992. Describes how to build e-mail applications and programs that use sockets and data streams under Unix, from the ground up.

[19] Cosmopolitan Virtual Makeover (Win/Mac), by SegaSoft, 1997. An interactive editor for augmenting and modifying the digitized appearance of a still 2D photo of the user. Uses a library of formatting information, including new hairstyles, lipstick, and eye shadow. Has no distribution channel for communication; runs on a single computer. All enhancements are performed manually.

[20] Adobe Premiere 5.0, by Adobe, 1998. Provides a good example of a current state-of-the-art movie editor for QuickTime movies. Demonstrates the basic and advanced features and commands required for a movie editor.

[21] Poser™ 3.0, by MetaCreations, 1998. Provides a good example of a current state-of-the-art joint- and parameter-driven humanoid body and face animation program. See Appendix 1 for the essential-information framework used by this system. A user interacts with a pose editor program and defines a series of static pose keyframes by using dials and typed numbers to specify joint angles and facial/body parameters. These can then be animated, by generating a series of static frames over a few minutes, which can then be saved to a QuickTime movie. No attempt is made to acquire the actual appearance or pose of the user. The system usually runs on only one computer; however, there is a facility for saving files of essential information, which can then in theory be e-mailed or distributed to another person also owning the Poser system for viewing. This system demonstrates that using essential information to generate presentations of detailed bodies and faces is well within the state of the art.

[22] Animated Conversation: Rule-based Generation of Facial Expressions, Gesture & Spoken Intonation for Multiple Conversational Agents, by Cassell, Pelachaud, & Badler, et al, Siggraph 1994, pp. 413-420. Describes a research system where two computer-animated human robots driven by artificial intelligences have a conversation with each other inside the same computer. The animated robots have fully-featured faces, arms, legs, and hands, and communicate using synthesized speech, facial expressions, bodily expressions, eye gaze, and hand gestures. All of the communications are synthesized inside the same computer by using essential information and formatting information. The system does not attempt to communicate with an actual human, nor does the system acquire the image of any actual users or environments. No distribution channel is used to communicate the essential information to a remote presentation system.

[23] A Structural Model of the Human Face, by Platt, U. Penn PhD thesis, 1985. Describes in explicit detail how to build an animated representation model of a fully-expressive human face. Uses the FACS structure for its essential-information paradigm. Discusses actual computer implementation of the FACS actions in full detail. Does not acquire images of users nor use a distribution channel, however, in an appendix Platt discusses the theory of how images could be acquired and used to represent a (unenhanced) face looking exactly like the user. The system was designed to have as output an image-modified (morphed) presentation of a single photograph of the user as taken head-on; however, this was never implemented in the thesis. The skeleton faces shown in the thesis are depictions of the internal representation designed for use in morphing. Did not discuss morphing between multiple images nor use of augmentations or costumes to enhance the presentation of the user into something different.

[24] Facial Action Coding System["FACS"], by Ekman and Friesen, Consulting Psychologists Press, Inc., 1978. Attempts to describe and codify every single atomic action that a human face can make. See Appendix 2 for a listing of these. Includes 135 photographs of faces performing different actions. Is used as the basis for most facial essential-information frameworks in use in research labs today.

[25] A Supervisory Collision-Avoidance System for Robot Controllers, by Myers, Carnegie-Mellon EE Master's Thesis, 1981, condensed and reprinted in Robotics Research and Advanced Applications, ed. by Book, The American Society of Mechanical Engineers, 1982, pp. 225-232. Demonstrates using essential information (joint-rotation vectors) to create an animated presentation of a moving robot, surrounding props, and its environment, using a changeable library of formatting information (models of the robot and environment to be used). Does not attempt to acquire images of human users.

[26] Robust Real-Time Face Tracking and Gesture Recognition, by Heinzmann and Zelinsky, IJCAI-97, pp. 1525-1530. Demonstrates that acquisition of the essential information describing a face and head is well within the state of the art. Uses custom algorithms on an off-the-shelf Fujitsu MEP tracking vision system. Demonstrates successful tracking and acquisition of the positions of the eyes, eyebrows, the shape of the mouth, gaze direction, and head orientation, robustly and at real-time frame rates. Puts the features together into temporal facial gestures such as "blink", "wink", "nod yes", "look down", etc. The system was designed as an input for a robotic assistant for paraplegics. It does not attempt to use a presentation to communicate to a viewer.

[27] Interfacing Sound Stream Segregation to Automatic Speech Recognition—Preliminary Results on Listening to Several Sounds Simultaneously, by Okuno et al., AAAI 96, pp. 1082-1089. Demonstrates preliminary results on abstracting the speech sounds of two users from a noisy environment. Then uses the speech sounds for automatic speech recognition of words. Does not transmit the words or the essential information in the speech sounds elsewhere to be presented for a viewer.

[28] Recognizing and interpreting gestures on a mobile robot, by Kortenkamp et al., AAAI 96, pp 915-921. Demonstrates tracking the head, shoulder, and arm of a user standing away from a mobile robot against a natural background. Abstracts essential information describing the arm, and recognizes poses and arm gestures.

[29] Approximate World Models: Incorporating Qualitative and Linguistic Information into Vision Systems, by Pinhanez and Bobick, AAAI 96, pp. 1116-1123. This system uses a world model of a TV chef and his bowls and food to track the chef's head, body, arms, and hands, and the objects in his environment, for output to robot studio-camera motions. The communicated signal is the video stream; no attempt is made at communicating the essential information, composed of the world models used by the tracking system, for use by a viewer.

[30] CVideo-Mail 1.2, by Cubic VideoComm Inc., 1997. A PC-based video/audio e-mail system. Sends movies attached to e-mail, which can be played by the receiving viewer. Movies are compressed literal images of the scene, and do not use essential information nor support enhancement changes.

[31] CineMail, by Baraka IntraCom, 1998. A PC-based video/audio e-mail system. Sends movies attached to e-mail, which can be played by the receiving viewer. Movies are compressed literal images of the scene, and do not use essential information nor support enhancement changes.

[32] Videogram Creator, bundled in QuickVideo Transport, by Alaris, 1998. A PC-based video/audio/text e-mail system. Sends movies attached to e-mail, which can be played by the receiving viewer. Movies are compressed literal images of the scene, and do not use essential information nor support enhancement changes.

[33] VDOPhone Direct by VDONet, bundled in QuickVideo Transport by Alaris, 1998. A PC-based videophone system. Videophone images are compressed literal images of the scene, and do not use essential information nor support enhancement changes.

[34] InfoView by InnoMedia, 1999. A TV set-top videophone that uses the POTS telephone network as its distribution channel. Runs at about 5 to 10 frames per second on a 352x288 or 176x144 image. Videophone images are compressed literal images of the scene, and do not use essential information nor support enhancement changes

SUMMARY OF THE INVENTION

A Fantasy Videophone supports communication by allowing a presentation of a *scene* (composed of a *user* plus *environment*) to be perceived by a *viewer*. The invention is composed of three main parts: the *imaging system*, the *distribution system*, and the *presentation system*. A *library* of presentation-construction *formatting information*, including cyberspace environments, avatar costumes, and voice fonts, is also typically used. The imaging system perceives the user's scene and abstracts *essential information* describing the user's *sensory appearance* (along with that of the environment). This appearance is primarily visual and auditory, although other sensory modalities (e.g. touch, taste, balance, smell, etc.) are possible. The distribution system transmits this information from the user's locale to the viewer's locale. And, the presentation system uses the essential information and the formatting information to construct a *presentation* of the scene's appearance for the viewer to perceive. The library of presentation-construction formatting information may be employed at one or more places in the system; it contributes information that is used, along with the abstracted essential information, to create the presentation for the viewer. Various parts of the entire communication system can be broken out separately as coherent components to be protected as well, including a *Fantasy Video Sender* including an imaging system, a *Fantasy Video Receiver* including a presentation system, and a *Fantasy Videophone Station* including both a Fantasy

Video Sender and a Fantasy Video Receiver. A key feature of the Fantasy Videophone invention is that the presentation representing the user plus environment does not have to attempt to be faithful to reality, but can rather be changed, that is, enhanced, in a number of different ways. Changes can be classified as *replacements*, *augmentations*, *modifications*, *deletions*, *filterings*, *overrides*, *repositionings*, *restagings*, or *combinations*. These changes support various results that are useful and surprising. The basic invention, comprising a system for a one-way point-to-point communication channel, can be used as a building block to construct various useful derived embodiments such as a two-way Fantasy Videophone; Fantasy Videophones with various connection topologies between one or more users, central servers, and one or more viewers; Sender or Receiver embodiments on various hardware configurations such as cell phones, wristwatch phones, wearable computers, wall panels, and PlayStation™-style set-top boxes, etc.; few-to-millions multicast "live" Fantasy Videophone TV-station configurations; and delayed-time systems such as Fantasy Videophone E-mail and "taped" Fantasy Videophone TV-station/receiver configurations.

Objects, advantages, and features include:

(a) To provide a Fantasy Videophone system that allows ordinary people to change the presentation of their actual environment over the Fantasy Videophone by replacing it with the photograph, scene, 3D model, painting, or other background of their choice; by augmenting it with the addition of new props, rooms, cars, or other creations of their choice; or by enhancing it by modifying, filtering, or removing objects that are portions of the environment;

(b) to provide a Fantasy Videophone system that allows ordinary people to change the presentation of themselves over the Fantasy Videophone by **replacing** their presentation with the costume or appearance of their choice;

(c) to provide a system that allows users to change the presentation of themselves over the Fantasy Videophone by **augmenting** their appearance with such props as horns, haloes, hats, extra armor, mustaches and beards, fangs, extra clothing, hair for bald users, etc.;

(d) to provide a system that allows users to change the presentation of themselves over the Fantasy Videophone by enhancing their appearance through such actions as modifying their apparent gender, modifying their age, modifying their race or apparent social background, deleting scars, tattoos, and blemishes, changing their hair color, or other desired enhancements;

(e) to provide a Fantasy Videophone system that allows ordinary users to change the appearance of their remote presentation by changing the apparent camera position and angle, lens zoom viewing angle, focus, depth of field, lens filters, lighting, colors, patterns, and other parameters of the scene's presentation;

(f) to provide a Fantasy Videophone system that allows ordinary people to change the remote presentation of their voice characteristics by replacing, augmenting, or enhancing the clarity, relative

volume, tonal quality, apparent age, size, gender, and identity of the speaker and the background environmental sounds, as part of the usage of a Fantasy Videophone;

(g) to provide a system that allows ordinary users to easily drive, in a single- or multi-user virtual world, avatars that reflect the facial expressions, speech, head and bodily movements of the user;

(h) to provide a system that allows users the capability of creating their own Fantasy Videophone "television programs" or "movies" that can be recorded, edited, and multi-cast to one or more viewers;

(i) to provide a Fantasy Videophone system that allows ordinary people to record Fantasy Video email, edit it, and send it to viewers;

(j) to provide a Fantasy Videophone system that uses significantly less bandwidth for communicating presentation information than current videophone methods;

Additional objects and advantages include:

The user can enter into a Fantasy Videophone conversation without revealing the user's actual appearance, gender, age, or ethnic background. This will allow users to maintain their anonymity while using the next generation of communication devices. Studies have shown that one of the reasons why email is so popular is because users like to maintain their anonymity. The Fantasy Videophone supports this feature.

The actual lighting for a user sitting in front of a computer screen is of poor quality. Since the screen is typically white, the user's face is floodlit with white light from the front, which typically washes out any other light in the room. This results in a flat and ugly image when transmitting the user's actual appearance on an ordinary videophone. However, the Fantasy Videophone can correct this problem by enhancing the appearance of the lighting applied to the user's facial presentation. The user can be made to look as if the user were calling from a perfectly-lit photography studio.

Similarly, many cameras have problems with incandescent or fluorescent lighting. The environment can come out with a yellow cast or a blue cast with normal videophones. The Fantasy Videophone can correct this by enhancing the appearance of the environment with a filter. Indeed, it is not even necessary to maintain natural lighting; the Fantasy Videophone can just as easily insert a cloud layer or a sunset in the back corner of an office conference room, put a starry black constellation-filled night sky out the window, run a laser show on a side wall, or even put the green atmosphere of Saturn, complete with rings, in the air around the user. Subtly natural enhancements or wild special effects are equally possible with various embodiments of the present invention.

A major problem with using a videophone that is mounted in a wristwatch, in a cellular phone, or in a handheld computer is that the distance of the camera from the user causes noticeable distortion in the regular, unchanged image of the viewer. In order to fit the entire viewer's face into the camera frame, the

camera lens must be wide-angle, causing barrel distortion in the regular image. In addition, because the user's nose is significantly proportionately closer to the lens than the rest of the face when the camera is held at wristwatch- or handheld-range, the user's nose looks extra large in the regular image and the rest of the face looks pushed back. Although useful for special effects, such appearance is typically considered ugly by most people for everyday communication. Because the Fantasy Videophone is not restricted to constructing the presentation with the apparent camera in the same place as the actual camera, the Fantasy Videophone can construct a presentation such that the user looks as if the user were being image-captured at a medium-far or far distance from the camera, and at an appropriate angle. This eliminates the distortion and results in a pleasing presentation.

A similar problem is found with wall-mounted videophones, when the camera is placed above, to one side of, or below the display screen that the user is facing head-on. In current art, this results in an unenhanced image that appears as if the viewer is looking down on, off to one side at, or up at the user at an unnatural angle. The Fantasy Videophone can enhance the image of the user by presenting it as if it were being taken by a camera placed behind the user's viewscreen, thereby allowing natural eye contact and relative head heights between the presentation of the user and the viewer. Alternatively, users or viewers may control the presentation to raise or lower the presentation's apparent head height or camera angle, in order to obtain feelings of superiority or inferiority.

The presentation can also be adjusted for focus in an arbitrary manner, unlike the actual image. Because the presentation is constructed, both objects that are near and far can be in perfect focus or can be presented out of focus in an arbitrary manner.

The focus modification can be used during interactive selections to choose which object in the presentation is to be the topic. This primitive capability can be used as a building block for handling interactions in a manner similar to the current paradigm of selecting an object in the computer by pointing a mouse cursor to it, double-clicking, and having the object turn gray or roll open. Other selection modifications include making the chosen object appear as if it were emitting light, change color, become encircled with a rainbow, become transparent, maintain solidity while the rest of the scene becomes transparent, unroll like a scroll, display a pop-up sign or menu, display a pie-chart, etc.

The viewer may control the appearance change by selecting an appearance for the user and for the user's environment according to the tastes of the viewer. For example, this would allow a viewer to add points onto the top of the image of the viewer's boss's hair, and to insert flames into the background, every time the boss called on the Fantasy Videophone. This could be customized per incoming caller.

The presentation of the user's body but not the user's face may be changed. This can result in presentations of talking trees and celery stalks with faces, similar to customary American elementary school plays. Similarly, the faces of the users may be changed but not their bodies, resulting in a masquerade party, the ability to impersonate a celebrity, or anthropomorphic talking wolves, etc.

The user may replace the background with a photograph of the same background that has been cleaned up. This photograph may be augmented or enhanced, e.g. by shifting the sunlight in the photo to correspond with the current time of day or with the outside weather automatically taken from a report.

This allows the viewer to always see a clean room, even when the actual room is messy. The apparent time of day can also be significantly modified to correspond with that of viewers in other time zones or on the other side of the world. In multiple-viewer systems, this can take place differently for each viewer in an appropriate manner.

A nonintrusive input system such as a video camera is easy and straightforward for the ordinary consumer to use, and does not require advanced expertise as the digitizing suits or glued reflective-marker systems do in order to be used in everyday life. A nonintrusive input system such as a video camera is inexpensive and may easily be bought by the ordinary consumer. A nonintrusive input system will typically require no significant setup time; it can simply be turned on and used. A nonintrusive input system will typically not require the use of an expensive and dedicated blue-screen room.

The long-awaited capability of being able to call on a video telephone while still preserving anonymity, or while wearing a virtual costume, results in a powerful synergy that will create important non-obvious ramifications significantly more powerful than those of its components taken together. Here are presented uses and ramifications that can be immediately projected. The population will no doubt create other uses that cannot be predicted yet.

The present invention may be used for such things as interpersonal communication, group communication, entertainment, game-playing, news, advertising, promotion, information retrieval, socializing, amusement, business, etc.

Allowing nonintrusive instantaneous expression-based control of avatars will result in a qualitatively significantly different experience for users. Because control is instantaneous, users will come to extend their morphic field over the avatars [in a manner similar to that done for eyeglasses, prosthetic limbs, and automobiles today] and feel like the avatar actually IS them. This experience will feel quite different from current avatar usage--almost as if the user has grown another arm.

Instinctive instantaneous control of avatars will open avatar cyberspace up to naïve users and the common citizen. Instantaneous control will also allow qualitatively more intense interactions between avatars.

A viewer can notify the Fantasy Videophone as to which physical types of person the viewer finds most attractive. Then the Fantasy Videophone can enhance the appearance of incoming calls from telephone operators, telesales people, and customer support people by combining the actual appearance of the caller with one of the ideal profiles in a randomly chosen manner. This enhances the quality of life of a viewer by ensuring that the viewer is always presented with callers the viewer finds attractive. It will also

be the birth of clearinghouse services that will store user-preference information and pass it on to telemarketers for a fee.

If a mother is not available, a remote baby sitter or day-care nurse can call young children over the Fantasy Videophone using a virtual costume of their mother, and thus calm them.

The communications link does not have to be long-distance. It can just as easily be used for house and apartment intercoms. A Fantasy Videophone link can be set up between the front doorbell and interior of a dwelling. If a stranger comes to door, a frail widowed housewife can use the virtual costume of a surly linebacker, complete with voice, and thus get rid of unwanted visitors.

Negotiations between widely differing national cultures can be eased if a user takes on the virtual costume of his or her appropriate counterpart in the other culture. For instance, when calling to a military dictator, an American negotiator can be given a chest full of medals in the Fantasy Videophone to enhance his status and make the dictator feel as if he was talking with an equal. Similarly, when calling to a traditionally formal Islamic country, an American woman can be presented as wearing a facial veil and body covering, whereas her Islamic friend can show herself in jeans and a T-shirt to the American.

Not merely the static appearance, but also the dynamic actions of a user can be filtered and cleaned up. Clumsy users can be made to move like dancers using the Fantasy Videophone. International calls between cultures with different body language, such as Vietnam and Italy, can be adjusted appropriately for the viewer. The Italian viewer can see wide and expansive gestures, while the Vietnamese viewer can watch quiet, restrained actions. This will eventually contribute towards easing international relations.

The virtual costume that is presented to the viewer is not limited to humanoid proportions. The user's appearance can just as easily be presented as an arbitrary object or abstract sculpture. For instance, a user could easily become a lung, a gliding seagull, a sunflower, or a DNA molecule. This would allow teachers to record performances or give live performances to students while teaching anatomy, literature, botany, or biology.

It is significant to point out, in the case of vast consensual cyberspaces, that large groups of viewers do not have to be instantiated in a scene when watching online performers. This allows 1000 viewers to all have the same single front row seat, without having to be worried by being blocked by the back of a head. Only the performers are shown to each viewer; views of the neighboring viewing masses are optional. This will allow 1000 students to stand in front of the Declaration and watch the signing performance all at the same time.

Because the Fantasy Videophone allows a user to take on whatever virtual costume the user chooses, this will facilitate storytelling immensely, both at the family and professional levels. Fathers and mothers can create beautiful stories and act them out in video for their children. Professional storytellers

will be able to use the Fantasy Videophone as a tool for generating work. Improv comedy artists and performance artists will be able to use the Fantasy Videophone to conduct paid performances, without requiring the rental of a theater.

Embodiments of the present invention will enable a new form of dance. Dancers will no longer be limited to a human shape, but will be able to take on such forms as a flock of birds, a group of leaves falling from a tree, a clump of clouds, a Christmas tree, or a group of dancing fairy lights from a fen.

Users will be able to custom-design their own cyber-homes and cyber-bodies. People and companies will be known by how creatively they construct and coordinate their presentations. A new form of popular art and culture will arise.

Teleministries will also benefit significantly from the Fantasy Videophone. No longer will broadcast preaching programs be restricted to evangelists with the hundreds of thousands of dollars required to buy a television program. Any local evangelist with a list of contact Videophone numbers will be able to create his or her own preaching program. Evangelists can use local talent to act out scenes from the holy book. Ambitious evangelists can use appearance enhancement and augmentation techniques to give themselves or their choir feathery wings, sunbeam spotlights, or golden haloes. Widespread broadcasting of Bible stories will serve to give the population a shared cultural identity, which will help combat the fragmentation caused by the deterioration of the nuclear family and its values. The dramatization and broadcast of scenes from other holy books will serve to combat ignorance and cultural intolerance in America. All this is made possible by the real-time, interactive, expressive characteristics of communication supported by the Fantasy Videophone.

Embodiments of the present invention will become an important part of culture. Right now, people talk with each other over the telephone. But, they have to wait until they can meet face-to-face in order to date and perform activities together. It is currently impossible for couples separated and living in separate states or countries to be able to meet and do things together without paying for an airplane ticket. With the Fantasy Videophone, users can create their own persona costumes and then go to central cyber bars or parties for meeting new friends. Industries will set up around providing quality meeting places and amusement parks for Fantasy Videophone users. Since the facial expressions and bodily gestures of the avatars will be controlled in real-time based on the actual expressions of the users, the users' experience will be an order of magnitude better and more immediate than the current non-real-time style of avatar control. A boy and a girl will be able to take a stroll together through a virtual flower garden. Couples will be able to date, even when they are remotely situated.

The Fantasy Videophone will significantly contribute to telecommuting. Because the system can work over very low bandwidths, there is no penalty for running a videoconference over plain old telephone service (POTS) lines. Anybody with simple Fantasy Videophone equipment can place a video call in to work, or set up a conference called between a numbers of colleagues.

In addition, the low bit volume required for storing a Fantasy Videophone email message will allow business users to send Fantasy Video email back and forth without having to delete every single message when it comes in because it is too large to keep on the user's hard disk, as is the case with current video mail technology.

The resulting increase in business traffic for the home office telecommuter will allow consultants to run entire businesses from their home without having to drive to work all week. As the nation shifts over to this mode of work, efficiencies will improve while air pollution goes down. Because the typical worker will no longer be forced to waste an hour and a half sitting in a pollution-generating automobile each workday, the population will have about 10 percent more free time to devote towards work improving the GNP or towards their personal lives. This is a significant amount. Widespread use of the Fantasy Videophone will contribute to bringing this situation about.

People who are drunk or hung-over can call home and look lively to their spouses. Hospital patients can similarly place Fantasy Videophone calls and present themselves as looking well, inside a normal environment. People with physical handicaps (disfigurements, missing limbs, tremors) and others who think that their appearance needs assistance will benefit from the device of the present invention.

Seniors and nursing home patients can flirt with other people they meet by appearing to be one-fourth of their actual age.

Young children can become the character of their choice and play together in a fantasyland. Play is enhanced by having the virtual costume characters reflect the facial expressions of the young users in real-time.

The Fantasy Videophone will allow sexual actors to perform for viewers without actually having to take their clothes off. The Videophone can replace the clothed bodies and faces of ordinary actors of both genders with nude virtual costumes of extraordinary appearance. Laws of physics and reality do not have to be respected in cyberspace; performances will be limited merely by the imaginations of their creators.

The Fantasy Videophone allows users to call in and visit a store while maintaining anonymity. The store can have an actual location, or it can exist merely in cyberspace. Anonymous shopping allows users to receive face-to-face personal attention from a store clerk while shopping for things they are embarrassed to be seen buying, such as ladies' underwear or contraceptives.

The Fantasy Videophone will also enhance sensitive conversations that benefit from facial information exchanges, such as anonymous drug counseling, family violence prevention, or suicide counseling. Fearful callers will feel more comfortable using a Fantasy Videophone than using a regular videophone.

Viewers can choose, on Fantasy Video TV, which actors' appearances they want to play which part. For instance, a viewer can choose to watch a James Bond presentation with Kermit the Frog playing the part of the hero, or have Raquel Welch play Hamlet. Viewers can even put their own appearances and those of their family into movies they watch.

It is possible to appear to be at home even when the user is calling from an arbitrary location. With caller ID, an appropriate environment can be selected and set up automatically, even when the user is accepting a call the user did not originate. For instance, if the user has a nosy aunt, the user can accept calls from the aunt at her boyfriend's house and still appear to be alone in a dorm room.

The filters and formatting information used do not have to be photorealistic. A user can be presented as a cartoon or animation drawing in a cartoon land, or as a pen-and-ink drawing, or as a sumi-e or impressionistic painting, or even as a cloud of lines or points.

A stereo 3D version of the Fantasy Videophone system can use two or more cameras or microphones for input. If the presentation device is capable of displaying in 3D stereo, the system can give the viewer a three-dimensional presentation. With the proper essential information and formatting information, the system can present a stereo 3D presentation even when the user's imaging system is two-dimensional. This can be done if three-dimensional avatar costumes and a three-dimensional environment are used, or even if two separate views are generated on two-dimensional morphed information, etc.

One of the most significant advantages of the Fantasy Videophone system is the incredible compression rates achievable from transmitting only essential information. For instance, say that a standard color-mapped video frame is 640 x 480 x 8 bits, or 307,200 bytes of information. If a system is set up so that there are 100 single-byte parameters that describe all the angles of the user's body, 100 single-byte parameters that describe all the positions of the user's facial features, 12 four-byte floating-point parameters that describe the suggested absolute global positions and orientations of the virtual camera and the user, and 59 bytes left over for extra information, this comes to 307 bytes per frame, resulting in a 1:1000 compression rate. Further care, such as transmitting only the changes in moving joints, can drop this extravagant rate to 60 bytes per frame or lower. This translates to 480 bits per frame, or at 30 frames per second, to a rate of 14,400 bits per second. Today's advanced PCs, using standard off-the-shelf 3D accelerator cards, are fast enough to generate simple scenes at 30 frames per second; tomorrow's PCs will have no problem with complex scenes. Thus an appropriately-configured Fantasy Videophone will be able to demonstrate full-frame 640x480 presentation information at 30 frames per second using information transmitted over a 14.4 Kbaud modem connection. Similarly-impressive results are achievable using voice fonts for sound compression.

Note that, since only the essential information is transmitted and the presentation is constructed for the viewer, the presentation system on the viewer's side can have a presentation device (e.g., screen or speaker system) of arbitrary resolution. For instance, the user can be using a tiny camera with only 8 bits

of color at 320x240 resolution, along with a 11 KHz sound input system; however, due to the advantages of this invention, the viewer can be watching on a 1080x1920 HDTV with six channels of 44.1 KHz sound. When avatar costumes, synthetic environments and voice fonts are used to replace the user's appearances, the resulting presentation can be generated to any arbitrary level of detail. Even when the user's appearances are augmented or modified, the system can still use interpolation and advanced entropy-based methods that integrate over time to construct a presentation that has more resolution than the imaging system acquires.

From the description above, a number of advantages of the Fantasy Videophone system become apparent:

(a) A user can choose to project an idealized view of the actual background instead of the current view. The user can record the background with all of the mess cleaned up and with high-status props (orchids, gold pens) added. The view can be automatically adjusted based on the time of day and the weather conditions.

(b) A user can choose to project a personally-chosen view of a different home or office, such as from "Lifestyles of the Rich and Famous", which shows off the temperament of the user.

(c) A user can choose to project other arbitrary scenes

(d) A user can choose to project an arbitrary computer-generated cyberspace world as the environment. This could be personal or consensual.

(e) A user can choose to project any other arbitrary thing for use as the background. This could include children's drawings, representations of favorite pets, famous paintings, magazine articles, text, whiteboard plans and scribbles, movies, abstract computer graphics, etc.

(f) Nothing in the invention restricts the system to using 2D information. The system can be constructed with two cameras for input, and a 3D display system for output, and can thus be used to replace, augment, or enhance 3D representations of the user.

The present invention includes method and apparatus for providing aVideo Circuit comprising an information transfer device enabled to allow at least one user to send a communication to at least one viewer; said information transfer device comprising an imaging system, one or more distribution channels, and a presentation system; said imaging system being enabled to acquire sensory information from said user and from the environment of said user; said sensory information being capable of being subjected to at least one enhancement by said imaging system; said enhancement being at least one of the following: a change, a replacement, an augmentation, a modification, a re-texturing, a cleaning up of said sensory information, a deletion, a filtering, an override, a reposition, a re-staging; said distribution channels being enabled to allow a communication from said user to said viewer; said sensory information representing at least one of the following: the appearance, sound, motion, and characteristics of said user; the appearance,

sound, motion, and characteristics of said environment of said user; said user controlling said enhancement of said sensory information from said user and said environment.

The present invention includes method and apparatus for providing an Image Information Representation Subsystem comprising a means for accepting digitized sensory images of the scene, a means for abstracting the essential information describing the appearances of the user(s) and/or their environment(s) composing the scene from the digitized sensory images, a means of representing this essential information, a means for making the represented essential information available for use or distribution. The present invention includes method and apparatus for making the represented essential information available for use or distribution. The present invention includes method and apparatus for abstracting the essential information describing the appearances of the user(s) and/or their environment(s). The present invention includes method and apparatus for representing this essential information.

The present invention includes method and apparatus for providing an imaging system comprising: a) the Image Information Representation Subsystem; one or more Image Acquisition Device(s) and one or more Means for Digitizing Images enable to take the sensory images from the Image Acquisition Device(s) or Means for Acquiring Images and convert them into a digital format for use by the Image Information Representation Subsystem.

The present invention includes method and apparatus for providing a Video Sender, comprising:

an Imaging System; a means of making the essential information available to a distribution channel whereby the one or more users, when using the Video Sender, can send sensory appearance essential information to the distribution channel or allow the channel to take the information, and whereby the one or more users can participate in the sending portion of a Video Sender conversation.

The present invention includes method and apparatus for providing a Presentation Construction Subsystem comprising a) a means for accepting essential information describing the scene, b) a means for creating sensory images from the essential information, c) a means for making the created sensory images available for use or presentation. The present invention includes method and apparatus for providingt he Presentation Construction Subsystem one or more Presentation Device(s) that distribute said sensory images to a viewer. The present invention includes method and apparatus for providing a Video Receiver comprising the Presentation System; a means of accepting essential information from a distribution channel whereby at least one viewer can view optionally changed presentations of at least one user and/or the users' environments based on said essential information being received from the distribution channel, and whereby the one or more viewers can watch Fantasy Video Movies or Fantasy Video TV Shows, or view the receiving portion of a Fantasy Videophone conversation.

The present invention includes method and apparatus for providing a Videophone Station, comprising the Video Receiver; of claim 7; Video Sender of claim 6; whereby the users can both send essential information and view presentations of one or more other users and their environments

The present invention includes method and apparatus for providing a Videophone Station, wherein one of said sensory image acquisition device(s) is at least one of the following: (a) a CCD camera, (b) a camcorder, (c) a stereo camera, (d) a microphone, (e) a stereo microphone, (f) a positional sensing device, (g) a geo-position sensing device, (h) a balance sensor, (i) an olfactory sensor, (j) a television camera, (k) a range sensor, (l) a spatial-occupancy sensor such as a laser light-stripe scanned array sensor, (m) a force-sensing joystick, (n) a biochemistry physiology sensor

The present invention includes method and apparatus for providing a Video Circuit, wherein one of said distribution channel(s) uses one or more of the following technologies: (a) the Internet; (b) a Local Area Network or Wide Area Network; (c) the telephone network; (d) computer tape; (e) the cellular telephone network; (f) CD-ROMs or DVD disks; (g) CDRs; (h) an Internet telephone; (i) cable typically used for cable TV; (j) fiber-optic cable; (k) radio waves, including the television broadcasting spectrum; (l) Web pages or FTP files

The present invention includes method and apparatus for providing a Presentation System wherein one of said presentation device(s) comprises one or more of the following devices: (a) a computer monitor; (b) a television; (c) a high-definition television; (d) a flat-panel display, such as is mounted on a wall; (e) a 3-D head-mounted display; (f) a system comprising a 3-D movie or computer monitor display, using lenticular lens gratings or LCD light-shutter devices in a flat panel or in viewers' glasses; (g) a hologram-making device; (h) a building-sized display sign; (i) a billboard; (j) a printer, color printer, photo printer, hologram film printer, hologram foil stamper, or color separation negative printer; (k) a picture-phone, screen phone, or videophone, including desktop and pay-phone styles; (l) a TV set-top device connected to a TV set or monitor, including cable boxes and family game computer systems; (m) a fax machine; (n) a cellular TV, picture-phone or videophone; (o) a wrist-watch TV or portable TV; (p) a wrist-watch picture-phone or videophone; (q) a laser-driven or N.C. router-based sculpting device, yielding output in wax, plastic, wood, metal, ice, or steel; (r) an LCD, dye, or plasma screen; (s) direct-to-magazine printers; (t) a laser-based device that projects an image directly onto the viewer's fovea; (u) a headset or wearable computer or fabric computer; (v) a window display on a vehicle such as an automobile, truck, bus, plane, helicopter, boat, tank, motorcycle, crane, etc.; (w) a neural transmitter that creates sensations directly in a viewer's body; (x) a computer-based movie projector or projection TV; (y) a hand-held game device; (z) a palmtop, laptop, notebook, or personal assistant computer; (aa) a screen display built into a seat or wall for use in the home, on airlines, inside cars, or in other vehicles; (bb) a computer monitor used in an arcade game or home computer game; (cc) a screen or speaker integrated with an appliance such as a refrigerator, toaster, pantry, or home-control system; (dd) any present or future device supporting a means for displaying a sensory presentation

The present invention includes method and apparatus for providing a Video Circuit, further including one or more libraries of formatting information describing specific methods and appearances for changing, enhancing, replacing, augmenting, modifying, retexturing, cleaning up, deleting, filtering,

overriding, repositioning, restaging, or changing in a combination of such enhancements the sensory appearance of such users and/or such users' environments, where such formatting information may include such forms as software "plug-ins" (external subroutines), 2D images, 3D images, solid models, morph spaces, cyberspace environments, avatar costumes, augmentation props, and voice fonts, among others, and where such formatting information is selected by a person or by a computer program, transferred into said presentation system, and used by said presentation system along with said essential information in creating said sensory appearances of the one or more users and the users' environments.

The present invention includes method and apparatus for providing a Presentation Construction Subsystem, wherein the presentation derived from the essential information and optional formatting information is constructed using one or more of the following technologies: (a) the essential information includes the positions and orientations of key parts of the users' bodies or the environments (b) the essential information includes the size and shape of key parts (c) the essential information includes joint angles, actuator parameters, and Costume Configuration Vectors for key joints, sets of joints, and configurations in the users' bodies or in the environments (d) the essential information includes routine calls in a graphics language that get interpreted or executed to help derive the presentation (e) the essential information includes codes for selecting display components from various sets, including such things as the identities or recommended identities of augmentations and replacements for key parts of the users' bodies or the environments, etc (f) the essential information includes points in a morph space, and the Presentation Construction Subsystem computes a regular morph or a perspective morph between different views to help construct the presentation (g) the essential information uses codes derived from the Facial Action Encoding System to help determine the presentation (h) the essential information includes a 3D model (i) the essential information includes a Camera Information packet that specifies the locations or characteristics of virtual cameras used in helping to construct the presentation (j) the essential information includes a Lighting Information packet that specifies the locations or characteristics of virtual light sources used in helping to construct the presentation (k) the essential information includes a Texture Information packet that specifies the locations or characteristics of textures used in helping to construct the presentation (l) the essential information includes a Literal Texture Information packet that specifies portions of one of the original acquired images to be used in helping to construct the presentation (m) the essential information includes combinations of the above, which are used in combination to help construct the presentation.

The present invention includes method and apparatus for providing a Videophone Cyberspace comprising a Fantasy Video Circuit; wherein said distribution channel further includes a third-party company that provides virtual costumes and/or virtual environments for one or more viewers, using one or more of the following technologies: (a) the Cyberspace is embodied by a computer, having a storage device, that is attached to the distribution channel and that acts as a server to send virtual costumes and/or virtual environments to said viewers (b) the Cyberspace is embodied by a CD-ROM or other storage device

containing virtual costumes and/or virtual environments that is sent to said viewers and read by a storage device reader attached to one or more of the viewers' presentation systems (c) the Cyberspace is downloaded over the Web or the Internet to local storage on one or more of the viewers' presentation systems (d) the viewer watches other users but is not embodied in the Cyberspace (e) the viewer is also a user that participates with other users in the Cyberspace (f) the user is the same as a viewer and interacts in a solitary manner with a remote program whereby users and viewers can participate in a solitary or consensual cyberspace, and users can control the speech, gestures, and expressions of their avatars in a natural manner.

The present invention includes method and apparatus for providing the Image Information Representation Subsystem wherein said Image Information Representation Subsystem uses sound images and contains: (a) means to abstract the essential information in the speech sounds of the user; (b) means to abstract a voice font that describes the voice characteristics of the user; (c) a means of representing this essential information

The present invention includes method and apparatus for providing the Presentation Construction Subsystem, wherein said Presentation Construction System uses essential information describing the speech of the user and formatting information having a voice font, and has (a) means to change the voice information by one or more of the following enhancements: replacing the voice font, augmenting the sound information with new information, modifying or filtering the existing sound information into something new, cleaning up or deleting parts of the sound information, overriding portions of the information with something different, repositioning the user's image in space, restaging the focus of the microphones, or a combination of such techniques; (b) a means to generate an internal sound image, using a voice font and said essential information whereby the true voice and sound environment of the user may be changed, enhanced, replaced, augmented, modified, retextured, cleaned up, deleted, filtered, overridden, repositioned, restaged, or changed in a combination of such enhancements, so that the viewer views (hears) an enhanced voice and sound environment, and so that the bandwidth requirements are relatively small

The present invention includes method and apparatus for providing a Fantasy Video email sending system comprising: a) the Imaging System of claim 4 wherein the system further includes a Fantasy Video email-sending engine having in addition: (a) an outgoing-message recording system that records image representation information from said image information representation subsystem into an outgoing e-mail message; (b) optionally, an outgoing-message storage buffer system into which said outgoing-message recording system records; (c) optionally, a re-record and review playback system that plays back a presentation of a previously-recorded outgoing message composed of said image representation information contained in said outgoing-message storage buffer system for user viewing, prompts for sending or deletion, and re-records the outgoing Video message if it is found to be unsuitable by the user; (d) an optional outgoing-message sending system that sends the e-mail to a distribution channel.

The present invention includes method and apparatus for providing a Fantasy-Video-email playing system comprising the Presentation System of claim 5 having in addition: (a) a means for playing back a message whereby the contents of a file associated with a Fantasy Video e-mail message are sent into said presentation construction subsystem of said presentation system; whereby a viewer can play back a Fantasy-Video email that the viewer has received, from inside a third-party web browser or other program that handles reception and storage of e-mail.

The present invention includes method and apparatus for providing a Fantasy Video email-receiving system comprising the Fantasy-Video-email playing system having in addition: (a) an optional incoming-message receiving system, which receives messages from a distribution channel; (b) a means for storing incoming messages, into which said incoming-message receiving system records; (c) an incoming-message playback system that takes a message from said mans for storing incoming messages and sends it into said presentation construction subsystem of said presentation system; (d) an optional viewer's message-selection system that selects which incoming messages to examine; whereby a viewer can work with a stand-alone system that is dedicated to the task of handling Fantasy-Video email.

The present invention includes method and apparatus for providing a Fantasy Videophone answering machine comprising the two-way Fantasy Videophone Station system wherein the system further includes on the called person's side: (a) a means for recording one or more Fantasy Videophone outgoing messages from the user's imaging system; (b) a means for outgoing-message storage, used by the outgoing-message recorder; (c) a means for playback of the recorded outgoing message from the storage, by sending it out the distribution channel to a viewer's presentation system when the viewer is calling in at a time the called person is absent or not picking up; (d) a means for controlling the functions of the answering machine that rings a signal for the called person, waits for call pickup, invokes the playback of said recorded outgoing message if the called person is absent, invokes the recording of the incoming message after the outgoing message playback has finished, and hangs up at an appropriate time after the caller is finished recording, after a time limit has been exceeded, or after the incoming-message storage is full; (e) a means for recording incoming messages; (f) a means for incoming-message storage, used by the incoming-message recorder; (g) a means for playback of the incoming message, usually from said incoming-message storage, by sending the incoming message to said local presentation system; (h) a means for choosing an incoming message to play back; (i) an optional means for review and re-recording of the outgoing message; (j) an optional means for selecting one out of many outgoing-message candidates, based on the caller ID, time of day, day of the week, holiday status, etc. (k) a means of "ringing" the called person by signaling that a call is coming in, used by the controller, whereby a Fantasy Videophone answering machine can ring the called person, wait for call pickup, select and play an appropriate Fantasy Videophone outgoing message if the called person is absent, recorded a Fantasy Videophone incoming message from the caller, and terminate the call.

The present invention includes method and apparatus for providing Fantasy Videophone TV broadcasting station with multi-track editing comprising the Imaging System wherein the system further includes on the users' side: (a) a means for recording the image representation information, called the "Fantasy Video Movie track", from said image information representation subsystem; (b) a means for Fantasy Video Movie storage, used by the recorder; (c) a means for playback and review of recorded Fantasy Video Movies from the storage, by sending them to a presentation system for viewing by an editor or director; (d) a means for editing the recorded Fantasy Video Movies, including for example methods for: (1) Laying down and merging a plurality of tracks together into a single Fantasy Videophone track; (2) Editing the existence, type, or intensity of changes in a track, including the choices of formatting information; (3) Shortening or lengthening a Fantasy Video Movie track; (4) Concatenating a plurality of tracks together into a single track; (5) Splicing a track into the middle of another track; (6) Adding special effects such as cross-fades; (7) Editing the voice fonts used, or the volume of sound from any particular actor; (e) an optional prompting system for the one or more users called "actors"; (f) a means for playback and "broadcasting" of the finished Fantasy Videophone Movie track by sending it out a distribution channel.

The present invention includes method and apparatus for providing Fantasy-Video Robot comprising a) an Image Information Representation Subsystem; an artificial intelligence program that plays at being a user while generating a Fantasy Video stream of essential information; whereby a computer can impersonate a person, and whereby a computer can present information such as run a help desk or an information kiosk in a remote fashion, and whereby one costly computer can simultaneously run multiple help desks that are presented on multiple inexpensive Fantasy Video Receivers, and whereby computer artificial intelligences can participate as actors in Fantasy Video TV/Movie dramas.

The present invention includes method and apparatus for providing a Fantasy-Video Robot, wherein said artificial intelligence also generates a stream of sound information that represents speech, which is also sent over a distribution channel, and which may or may not be encoded using Fantasy-Video essential information methods; whereby a computer artificial intelligence can communicate intelligible information to a viewer, such as in a help desk that speaks in multiple languages, and whereby a computer artificial intelligence can communicate unintelligible information to a viewer, for example with a robot Furby or Klingon character, and whereby a computer artificial intelligence can speak over a telephone, or whereby a computer artificial intelligence can speak over a Fantasy Video Circuit

The present invention includes method and apparatus for providing a Fantasy-Video Robot, being a Two-way Conversing Fantasy-Video Robot, wherein said artificial intelligence also includes means for accepting information over a distribution channel from a conversing viewer, and said artificial intelligence also includes means for responding to that information in an interactive conversing manner; wherein said means for accepting information over a distribution channel includes a means for accepting information derived from the speech of said conversing viewer; wherein the means for accepting information from a

non-colocated conversing viewer includes Fantasy-Video essential information derived from sensory images of a viewer, including such modalities as visual images or sound images

The present invention includes method and apparatus for providing Fantasy Videophone cyber bar, dating club, or amusement park, comprising a Fantasy Videophone Cyberspace system wherein the system provides extensive support for one or more users called "customers" and zero or more viewers who are not users, called "lurkers" or "spectators", and also provides one or more interesting environments

The present invention includes method and apparatus for providing the Presentation Construction Subsystem wherein the mathematical camera location used in generating the presentation of the user and the environment is under the control of the viewer or user or an automatic tracking program, and can be switched between a first-person viewpoint, an over-the-shoulder viewpoint, or zero or more remote-camera viewpoints, or can be restaged so that it is pointing at the presentation of the user from a reasonable distance and angle, and/or can have its virtual lens-angle adjusted

The present invention includes method and apparatus for providing Dual-Channel Fantasy Videophone system comprising one or more Fantasy Video Circuits arranged in a topological configuration such that some users are also viewers, in particular in one or more of the following topologies: (a) a "ring" topology, wherein each caller connects directly with one other user and with one other viewer, and user presentation information is relayed and passed around a ring of Fantasy Videophones; (b) a "mesh" topology, wherein each caller connects directly with all other callers and views them directly; (c) a "star" topology, wherein each caller connects directly with a central server that relays the information from each user to all viewers, and may add information itself; (d) A "broadcast" topology, wherein one user communicates in a one-way or two-way fashion with multiple viewers; (e) a "multi-to-many relay" topology, wherein a plurality of users communicate in a one-way or two-way fashion through a central broadcasting relay station with multiple viewers; (f) a "multi-to-many direct" topology, wherein a plurality of users communicate in a one-way or two-way fashion directly with multiple viewers; (g) a "network" topology, wherein various switches, routers, and repeaters are used to build and break dedicated or transitory connections, and a "yellow pages" network manager may be used to help locate called parties and establish connections; (h) other common topologies not listed here.

The present invention includes method and apparatus for providing the Imaging System wherein the image acquisition subsystem has multiple cameras trained on a single scene, which allow the imaging system to more easily acquire the 3-D locations of features in the scene.

The present invention includes method and apparatus for providing the Fantasy Video Sender of wherein additional channels of information, such as sound, video, or text, are also transmitted over one or more distribution channels without the benefit of Fantasy Video essential-information encoding, for example, a Fantasy Video Sender that also transmits unencoded sound over the same distribution channel or over a regular telephone circuit, or a Fantasy Video Sender that is also part of a regular TV broadcast, etc

The present invention includes method and apparatus for providing the Fantasy Videophone Station, wherein said Image Information Representation Subsystem and said Presentation Construction Subsystem are implemented on a PC computer or wherein said Image Information Representation Subsystem and said Presentation Construction Subsystem are implemented on a family game play station computer or wherein said Image Information Representation Subsystem and said Presentation Construction Subsystem are implemented on a wearable computer or in a wristwatch format

The present invention includes method and apparatus for providing the Presentation Construction Subsystem having in addition a multimedia presentation engine that is capable of showing linear or branching nonlinear interactive or non-interactive presentations of media including media such as one or more of the following: text, 2D images, 3D images, 2D movies and animation, 3D movies and animation, avatar animation, morphing animation, 3D model animation, DVD movies, sound, MIDI, or triggered events such as laser light shows and curtain openings

The present invention includes method and apparatus for providing the Presentation Construction Subsystem, having in addition an interactive or non-interactive display to a hypertext or hypermedia system that is capable of following links to hypertext or hypermedia information nodes called "pages", including technologies such as one or more of the following: (a) a World Wide Web browser, (b) a local Web browser, (c) a Gopher system, (d) FTP; wherein optionally the Receiver permits "hot spot" invocations to be areas or semantic parts of the Fantasy Video presentation, such as a user's avatar's eyes or hands, etc., so that the viewer can follow links by selecting different portions of the presentation of the users or the users' environments.

The present invention includes method and apparatus for providing the Presentation Construction Subsystem, having in addition a computer game that may be local or distributed over a network

Further objects and advantages of the invention will become apparent from a consideration of the drawings and ensuing description.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows an overview of the basic Fantasy Video Circuit design, wherein a viewer watches a sensory presentation of a scene composed of a user plus the user's environment over a video system having an imaging system, a distribution channel, and a presentation system.

Fig. 1B shows the Fantasy Video Sender, being the first main portion of a basic Fantasy Video Circuit, which is composed of an imaging system plus a means of making its resulting "essential information" available over a distribution channel.

Fig. 1C shows the complement to Fig. 1B, the Fantasy Video Receiver, being the second main portion of a basic Fantasy Video Circuit, which is composed of a means of accepting information from a distribution channel plus a presentation system that converts it into a sensory presentation for the viewer.

Fig. 1D contrasts a one-way Fantasy Video Circuit typically used with for a Fantasy Video TV embodiment with a two-way Fantasy Video Circuit typically used for a Fantasy Videophone. In the second case, each person both uses an imaging system and also views a presentation system.

Fig. 1E shows a Fantasy Videophone Station having both a Fantasy Video Sender and a Fantasy Video Receiver, which is a main portion of a two-way Fantasy Video Circuit. It is used by a person who both uses its imaging system and views its presentation system.

Fig. 1F emphasizes that an imaging system can use multiple sensory Image Acquisition Devices to acquire scenes, including such things as cameras to acquire visual images and microphones to acquire sound images. The Image Acquisition Devices can be used to acquire multiple 2D or 3D scenes.

Fig. 2 shows the first step in an embodiment of the Image Information Representation Subsystem that abstracts essential information by first extracting the sensory image of the user from the sensory image of the user's environment. Shown are input and output images for the extraction process for both video and sound images.

Fig. 3A shows a conceptual diagram of the main methods used by the preferred embodiment of the system based on actuator-variable values and robotics/3D computer graphics methods.

Fig. 3B shows a conceptual diagram of the main methods used by an alternative embodiment of the system based on perspective-morphing 2D images between different samples that define a "morph space" of possible images of poses.

Fig. 4 illustrates example classes of types of changes/enhancements that can be applied when constructing the Fantasy Video presentation.

Fig. 5 shows some of the features typically used to encode the essential information representing the facial configuration of the user.

Fig. 6 presents example distribution channel embodiments for point-to-point communication from a single user to a single viewer. Other examples are obvious.

Fig. 7 presents example distribution channel embodiments for different communication topologies between one or more users and one or more viewers.

Fig. 8A, Fig. 8B, and Fig. 8C present example presentation devices.

Figs 9A, 9B and 9C show different alternative embodiments for location of an optional library of presentation-construction formatting information having algorithms, filters, virtual costumes and virtual environments, etc. Fig. 9D shows an embodiment combining possibilities from the previous alternatives.

Fig. 10 shows an embodiment for a Fantasy Video TV Broadcasting Station with multitrack editing.

Fig. 10B shows an embodiment for a Fantasy Video Recorder.

Fig. 10C show an embodiment for a Fantasy Video Editor.

Fig. 11 shows an embodiment for a Fantasy Video Email system.

Fig. 12 shows an embodiment for a Fantasy Videophone Answering Machine.

Fig. 13 shows an embodiment for a Presentation-Construction Information Editor that can interactively create interesting virtual costumes and virtual environments for later use by the user.

Fig. 14 shows how the invention can correct camera-positioning and lens problems in a videophone image by restaging the shot by moving the position of the virtual camera and changing its parameters to make a clean presentation. The upper images show how the user actually appears to the imaging system in two common embodiments of the Image Acquisition Device; the lower image shows how the presentation system can restage this scene for the viewer.

Fig. 15A shows an embodiment of a Fantasy Video Circuit using set-top boxes that are computer game play stations being connected by a telephone/Internet communications cloud.

Fig. 15B shows an embodiment of a Fantasy Videophone Station that uses a set-top box that is a computer game play station. The drawing could also be an illustration of a Fantasy Video Sender with an extra TV, or a Fantasy Video Receiver with an extra camera.

Fig. 15C shows an embodiment of a Fantasy Videophone Station that uses a computer, a camera, and a computer monitor.

Fig. 16A shows an embodiment of a Fantasy Video Receiver that also has a multimedia presentation system, including such systems as a laser light show.

Fig. 16B shows an embodiment of a Fantasy Video Receiver that also has an interactive hypermedia system such as a browsing interface to a web or network of information.

Fig. 16C shows an embodiment of a Fantasy Video Receiver that also has an interactive computer game.

MODES FOR CARRYING OUT THE INVENTION

DEFINITIONS

User: a person sending presentation information over a Fantasy Video Circuit

Viewer: a person receiving presentation information from a Fantasy Video Circuit

Environment: the background, foreground, and surrounding objects that are around the user. Also the presentation of background, foreground, and surrounding objects that are presented in addition to the presentation of the user and that are viewed by the viewer

Scene: the user plus the environment

To Present: A verb meaning "to display", except that it covers all sensory modalities

Presentation: The sensory appearance of a scene, as presented for perception to the viewer

Appearance: An input to the senses. Includes visual, auditory, tactile, force, taste, olfactory, balance, distance, shape, and kinesthetic etc. modalities

Image: A perceived or captured appearance, especially as represented inside a computer. Includes all sensory modalities, such as visual images, sound images, force images, etc.. Images can be acquired for computer use, or they can be presented for the use of the viewer.

Sensory Image: Emphasizes the fact that an image can be in any sensory modality

Changing the appearance: Making a presentation of the user or environment that is slightly or largely different than the literal images of the user or environment acquired by the image acquisition device

Enhancing the appearance: Changing the appearance in a manner that has been specified by someone or by a program

Types of appearance changes or enhancements: replacement, augmentation, modification, cleaning up, deletion, filtering, overriding, repositioning, restaging, combination changes

Replacing the appearance: total substitution of a part or whole with something completely new

Augmenting the appearance: adding new parts or entities to the existing appearance

Modifying the appearance: changing existing parts so that they are different, while retaining their identities

Cleaning up the appearance: removing or deleting some features of parts, such as freckles or dirt

Deleting the appearance: removing or deleting parts or a whole, such as long hair, clothing, etc.

Filtering the appearance: changing non-part appearances such as color, etc. Typically 2D.

Overriding the appearance: Changing the movement characteristics of a part or whole. Includes making a body move differently, or in a different manner. Includes putting different emotions into sound conversations, or overriding the words spoken.

Repositioning the appearance: Moving parts of the scene relative to the virtual camera. Is one form of restaging.

Restaging the appearance: Changing the parameters of the virtual lighting and cameras, including camera position and orientation, lens angle, focal length, lighting color, lighting quality, and the existence of different virtual cameras and lights. Also includes repositioning parts of the scene.

Combination changes of the appearance: Using a combination of more than one of the above changes.

Essential Information: A vector in a state space that describes the essential content of what is going on in a scene. Can include such things as the joint angle of each of a user's joints, the components of an expression on a user's face, the position and orientation of a user, the pitch and loudness of a user's voice, the phonemes that a user is uttering, the emotions that a user is conveying, the identity of the user, the position of the user's face in a morph space, the position of movable objects such as curtains in the environment, etc. Can also include video patches that send over the actual appearance of the user's eyes, eyebrows, and mouth, or even literal images of the user abstracted from the background in basic systems. Will typically include a literal stream of the audio in all but very advanced systems. May often include a literal snapshot of the empty environment without the user, sent once at the beginning. Analogous to the unformatted ASCII content of a Web page's text.

Formatting Information: Extra information describing the appearance of the user and the environment that complements the essential information, and is necessary to use in order to reconstitute the essential information into some form of (typically enhanced) appearance. Can include such things as: 2D or 3D backdrop photos, of the actual environment or of something else; cyberspace environments; avatar bodies and faces; voice fonts; 2D, morph space, or 3D models of the actual appearance of the user; similar models of the appearance of other users; virtual environmental props and machines; local "bot" intelligences sufficient to run simple non-user characters or active parts of a cyberspace environment; augmentation props, such as wings, a long white beard, and a halo; colors and textures for filling out solid models; modification routines and specification information e.g. for making only the legs and the head of the user's appearance twice as long; deletion routines and specifications e.g. for removing only the hair and the legs of the user's appearance; filtering plug-ins e.g. for making the image of the user twirl, or be projected onto a mosaic grid; overriding routines and specification information e.g. for making the appearance of the user's legs dance, and making the user move like and have the body language of

an Italian person; repositioning routines and specifications e.g. for moving the user back from the virtual camera, or the virtual camera back from the user; restaging routines and specifications e.g. for changing the apparent lighting in the scene, the gross camera location, the angle of the lens, and the depth of field; and combination information, for multiple enhancements. Formatting information is analogous to the HTTP commands of a Web page that describe the appearance of the presentation of the ASCII content of the page's text.

Literal Image, or Literal Texture Image: For the most part, the "essential information" that is abstracted and transferred over the distribution channel involves a few variables that are selection codes or that describe particular configurations. In some cases, however, the essential information could involve using significant subsets of the image, where the subset is less than the entire image. In this case the usage involves simply sending these subsets across in a literal fashion (or perhaps encoded using a traditional compression algorithm) instead of encoding the semantics of what the image "means" or what is going on in the image. For instance, the essential information could be an image of only the user with the environment removed, or it could be images of only the user's eyes and mouth, as cut in a window directly from the digitized image acquired from the sensors. This is called a "literal image", and it applies especially to pictorial images. A "literal texture image" reinforces that the image in question is pictorial, and is used especially for images that will be used as "texture maps" to add photographic coloring to a 3D computer model or will be used in a morphing operation.

Virtual costume: the formatting information for an appearance change for a user. Especially used for a body and face replacement, possibly with a voice replacement

Voice font: The set of formatting information needed to specify the vocal characteristics of the auditory portion of a speaking user's presentation. A costume for the voice, used in a manner analogous to how letter fonts are used to present typography. When combined with the essential information of a speaker's utterance, allows the construction of a vocal presentation.

cyberspace: A set of virtual environments typically provided by a third-party company, usually along with a set of virtual costumes for the user to select from. A cyberspace is usually implemented with a central server, run by the third-party company, that acts as a relay between multiple users and viewers. A single presentation will usually combine representations of multiple users from multiple imaging systems into the same scene. Connections are typically two-way, with each user also viewing the communal scene.

avatar: A virtual costume, especially one used in a cyberspace.

stream: A reification of a temporal sequence of information, i.e. "a stream of information". Has the same technical meaning as in Unix.

station: A device that embodies the imaging system on the user's side, or the presentation system on the viewer's side, or both the imaging system and a local presentation system for a viewer using a two-way Fantasy Videophone.

view: To perceive a sensory presentation. Includes all sensory channels such as visual, auditory, tactile, and force channels, etc.

position: A technical term meaning the [x, y, z] or [x, y] positional coordinates of a point or an object in some space. Position in 3-space has three degrees of freedom.

orientation: A technical term meaning the direction that something points in space. This can typically be represented using [roll, pitch, yaw], quaternions, or a 3x3 matrix. Orientation in 3-space also has three degrees of freedom.

location: A technical term meaning the position plus the orientation of an object. Locations can typically be represented by a 4x4 homogeneous matrix or a point plus a quaternion, etc. Location in 3-space has six degrees of freedom.

The inventor contemplates marketing the device of the present invention as a "Fantasy Videophone" and uses the term Fantasy throughout this application to further assist in identifying his inventions.

Fig. 1 shows the Fantasy Video Circuit that is the primary building block for the embodiments. One or more users 0001 send sensory information to one or more viewers 0005. The one or more users 0001 are positioned amidst their environment(s) 0002. The environment optionally includes elements behind, to the side, and in front of the user. A user 0001 might consider the environment 0002 to be ugly, and might want to change its appearance in the Fantasy Video Circuit. A scene 0004 consists of an environment 0002 plus any users 0001 in that environment 0002, as perceived by a single imaging system 0010. One imaging system 0010 is enabled to process one, or more than one, scenes 0004. The user 0001 is optionally absent, or there can be more than one user 0001 in such a scene 0004. An imaging system 0010 acquires sensory images of the user 0001 and the environment 0002. The imaging system 0010 has at least an image acquisition subsystem 0020 and an image information representation subsystem 0021; more complex embodiments, such as a Fantasy Video Email system or a Fantasy Videophone Answering Machine will have more components accompanying the imaging system 0010. The image acquisition subsystem 0020 has at least a means for acquiring sensory images or an image acquisition device 0023, coupled with an appropriate means for digitizing images 0024. The means for acquiring sensory images or image acquisition device 0023 accepts raw sensory information from the user 0001 and the environment 0002, and captures this raw sensory information in a usable form; typical examples are a CCD camera or a microphone. The means for digitizing images 0024 converts this usable form into a digital (numeric) format that can be used by digital processors. Typical examples of means for digitizing images 0024

include a sound ADC or a video capture board. In some embodiments the image acquisition device 0023 and the means for digitizing images 0024 will be embodied on the same device, for instance a digital microphone or a CCD camera that gives output directly in digital format. Thus, the imaging system 0010 uses the image acquisition subsystem 0020 to acquire visual, auditory, and other sensory images of any users 0001 and their environment(s) 0002 in a digital format. Then the image information representation subsystem 0021 takes these digital images, abstracts the essential information in these images and represents this essential information in a form that allows distribution. The image information representation subsystem 0021 will typically be embodied in a computer, a CPU, or a custom hardware processing chip. The essential information could be portions of the image itself, it could be an encoding of the contents of the image, or it could be an encoding of the information in the scene, such as the pose of the user. The means for digitizing images 0024 and the image information representation system 0021 can be conceptually grouped as the means for processing image 0025. Alternate embodiments include a computer with a digitizer card, or a custom chip for a wrist-watch video-phone that digitizes and processes the image. The representation of the resulting essential information is transmitted over a distribution channel 0011 to a presentation system 0012. The distribution channel 0011 is preferably a real-time communication medium, such as the telephone network; or it can be a non-real-time information transfer channel such as e-mail or printing a CD-ROM disk and selling it in stores for viewers to buy and use. The distribution channel 0011 carries the essential information to the presentation system 0012, which creates a sensory presentation 0060 for one or more viewers 0005 to perceive. The presentation system 0012 has at least a method for presentation construction 0030 that takes the essential information and creates a presentation internally, and a presentation device 0040 that takes this internal presentation and presents it as an (external) presentation 0060 for the viewer(s) 0005 to perceive. The presentation 0060 optionally features portions representing the environment 0052, and portions representing any users 0050, along with any overlay information, and other information, etc.

A key feature is that the presentation of the user 0050 and/or the presentation of the environment 0052 can be changed by the system. This allows the user(s) 0001 to change the appearance of the environment(s) 0002 and/or to change the appearance of the user(s) 0001.

As mentioned in the definitions, the user is the person transmitting the image, and the viewer is the person receiving the image. Sometimes viewers will also be users, and vice versa.

The Fantasy Video Circuit illustrated in Fig 1 is a system for communication. Although all parts are necessary in order to communicate, the major portions of the circuit, including the Fantasy Video Sender 0008 described in Fig. 1B, the distribution channel described in Fig. 6, and the Fantasy Video Receiver 0009 described in Fig. 1C, will typically be disjoint, heterogeneous parts. That is, a Fantasy Video Receiver will often be implemented using different technology from a Fantasy Video Sender. For

instance, a user sitting in front of a PC-based Fantasy Videophone could be having a conversation with a viewer using a cell-phone based Fantasy Videophone.

Fig. 1B shows the Fantasy Video Sender 0008 which includes the Imaging System 0010, and a means for making information available to the distribution channel, 0014. The Fantasy Video Sender 0008 uses the Imaging System 0010 to acquire image information from the scene(s) 0006 of the user(s) 0001 and their environment(s) 0002, and abstract this image information into an essential information representation inside the Imaging System. Then, the means for making information available to the distribution channel 0014 enables distribution to elsewhere by means of a distribution channel 0011.

Fig 1C shows the Fantasy Video Receiver 0009 which includes a means for accepting information from the distribution channel, 0017, and the Presentation System 0012. The Fantasy Video Receiver 0009 accepts information, including signaling, image essential information, command and control information, and other information, from a distribution channel 0011 by means of the means for accepting information from the distribution channel 0017. Then the Presentation System 0012 uses the Presentation Construction Subsystem 0030 to use this information, especially the essential information, to help in constructing the presentation 0060 of the user(s) 0050 plus their environment(s) 0052 for the viewer(s) 0005 to view, by means of the Presentation Device 0040. Note that the presentation 0060 is optionally an ongoing presentation that is displayed whether there is any essential information present or not; this is useful if the distribution channel is sporadic, lossy, or has unpredictable delay times.

Fig. 1D illustrates that two of the one-way basic building-block Fantasy Video Circuits from Fig 1 can be set up in opposite directions to support a two-way conversation. It contrasts a one-way Fantasy Video circuit with a two-way Fantasy Video circuit composed of two one-way circuits. The one-way Fantasy Video circuit is used in applications such as a Fantasy Video TV, in which the viewer 0005 sees a presentation of the user 0001, but the user 0001 does not see a presentation of the viewer 0005. The two-way circuit is used in applications such as a Fantasy Videophone, in which the caller is a person 0003b acting as both a user and a viewer, and the called person is also a person 0003 acting as both a user and a viewer--that is, each person can see the other person. In either case, an Imaging System 0010 acquires images of the user 0001, 0003, or 0003b, abstracts the essential information and sends it out along the distribution channel 0011, where it is made available to and used by the presentation system 0012 on the opposite side. Thus the viewer 0005 or person acting as a viewer 0003b, 0003, is enabled to view a presentation of the user 0001 or person-acting-as-a-user 0003, 0003b respectively. Some applications will also insert a presentation of the user 0001 into the scene that is fed back for the user 0001 to view, so the person can view a presentation of him or herself from a third-person view.

Note that the distribution channels in the two-way circuit may or may not use the same technology or the same channels. For example, one distribution channel could be embodied as a two-way circuit on a cell phone that is used simultaneously by both Fantasy Video Circuits. Conversely, for example, one

distribution channel in a two-way circuit is alternatively embodied using a cable TV medium, whereas the other channel is alternatively embodied using a telephone network.

The two-way configuration is not restricted to a point-to-point topology. Two-way circuits can compose multicast configurations, as will be discussed in Fig 7.

Fig 1E shows a Fantasy Videophone Station 0007 that is a device for one side of a system used for two-way communications as previously discussed in Fig 1D. The Fantasy Videophone Station 0007 has at least a Fantasy Video Sender 0008 and a Fantasy Video Receiver 0009. The users(s)-also-acting-as-viewer(s) 0003 can have their images acquired by the one or more Image Acquisition Device(s) 0023, and can also view presentations 0060 on the one or more presentation device(s) 0040, typically in a simultaneous manner. The Fantasy Video Sender 0008 uses the Imaging System 0010 to acquire sensory images of the user (possibly also acting as a viewer) 0003 by means of one or more of the Image Acquisition Devices 0023, digitize these images using the means for digitizing images 0024, and represent the essential information in these images using the image information representation subsystem 0021. Then the essential information and any necessary control information is passed out using the means for making information available to the distribution channel 0014. A distribution channel 0011 will carry this information to a viewer in a different locale. The Fantasy Videophone Station 0007 also uses the Fantasy Video Receiver 0009 having the means for accepting information from the distribution channel 0017 and the Presentation System 0012. The Fantasy Video Receiver 0009 accepts information from a distribution channel 0011 using the means for accepting information from the distribution channel 0017; this information is fed to the Presentation Construction Subsystem 0030 which uses it to help construct the presentation 0060 of the other-locale user(s) 0050 and their environment(s) 0052, which is displayed on the Presentation Device 0040 for viewing. In this manner, the user-also-acting-as-a-viewer 0003 is enabled to send information and is enabled to receive information.

Typically, the sending circuit and the receiving circuit will be attached to the same remote caller or group of callers. However, in certain rare cases, the user-acting-as-a-viewer 0003 might view one group of people while sending information to another group of people.

Fig. 1F reinforces the fact that a single Imaging System 0010 might have multiple Image Acquisition Devices 0023. Here we see a system with three camera Image Acquisition Devices 0023 and three microphone Image Acquisition Devices 0023. The cameras gather visual images, while the microphones gather sound images. These images are digitized by the Means for Digitizing Images 0024, which could be one digitizing card or several digitizing cards, and then used by the Image Information Representation Subsystem 0021 to help form a representation of the essential information in the scene. Other information used by the Image Information Representation Subsystem 0021 may include historical information, for example for use in Kalman Filters, and calibration information, along with any attention-selection information. The essential information and any required control or handshaking information is

then given to the Means for Making Information Available to the Distribution Channel 0014. The distribution channel 0011 takes this information and transports it elsewhere. The three cameras are used to form a more comprehensive acquisition of the scene of the user(s) 0001 and the environment(s) 0002; often they will be used to form a 3D model of the scene. The three microphones also are used to form a more comprehensive acquisition of the scene of the user(s) 0001 and the environment(s) 0002. Three or four microphones are alternately used by the Image Information Representation Subsystem 0021 to triangulate the source of a sound, and help separate user speech images from background environmental noises.

Fig. 2 illustrates alternate first steps in gathering the essential information. The Image Information Representation Subsystem 0021 accepts the original sensory image of the user plus environment (video version), 0260a. It then typically separates the extracted appearance of the user (video version), 0270a, from the extracted appearance of the environment (video version), 0280a. These may then be used separately for further processing. Typically the extracted appearance of the environment (video version) 0280a will be sent to a Fantasy Video Receiver 0009 at the beginning of a session, and kept and used by its Presentation System's (0012) Presentation Construction Subsystem 0030 in case it is needed, e.g. for presenting an augmented or replaced costume for the user inside a normal environment, or for repositioning or restaging the presentation of the user in the environment. However, the extracted appearance of the environment will typically be sent only once, or at a slow rate, e.g. once per minute. Whereas, the extracted appearance of the user (video version) 0270a may constitute the essential information itself in a simple system, and thus be sent as often as possible, or it may be used for further pose acquisition by the Image Information Representation Subsystem 0021, e.g. to extract essential information comprising head position, head orientation, eye blink, or literal images of the eyes and mouth only, etc. One straightforward method for separating the appearance of the user from the appearance of the environment is to ask the user to step out of range of the camera sensor(s) and record the image of the environment by itself. Then, when the user steps back in and is using the Fantasy Video Sender 0008, the standardized image of the environment is compared against the current original sensory image of the user plus the environment (video version) 0260a, and any pixels that do not match in both color and surrounding texture are proposed as candidates for the extracted image of the user only (video version) 0270a.

A similar process can be used for other sensory modalities such as sound, as is shown in the bottom half of Fig. 2. The Image Information Representation Subsystem 0021 accepts the original sensory image of the user plus environment (sound version), 0260b. It then may separate the extracted appearance of the user (sound version) 0270b, from the extracted appearance of the environment (sound version), 0280b. These may then be used separately for further processing. Often the extracted appearance of the environment (sound version) 0280b will simply be discarded, while the extracted appearance of the user (sound version) 0270b will have its essential information extracted, encoded, compressed, and sent across a distribution channel 0011 to a Fantasy Video Receiver 0009. Because the Image Information Representation Subsystem 0021 first typically extracts the extracted appearance of the user (sound version)

0270b from the original sensory image of the user plus environment (sound version) 0260b, the rest of the processing used by the Image Information Representation Subsystem 0021 to extract the essential information of the image is made a lot simpler.

Fig. 3A illustrates a typical paradigm for essential information and presentations based on solid models, world models, and 3D computer graphics. The image information representation subsystem 0021 of the Imaging System 0010 uses a software face- and body-tracker algorithm to track and acquire features of the user(s) 0001 and the environment(s) 0002. These features are abstracted into a model, and then the model is abstracted into essential information. This essential information is sent over the distribution channel 0011 to the Presentation System 0012 on the other side, where it is used to construct a presentation of the user 0050. In this paradigm, the essential information consists of all of the information needed to create a computer-graphics-based presentation of the user. First, at initialization time, it is necessary for the user 0001 or viewer 0005 to select a Costume Or Enhancement Selection Specification 0200 that will typically be a code or a series of commands that specify which Costume Model 0205 will be used, or the nature of the appearance enhancements to be employed. Figure 3A shows replacement using a simple blocks robot model. The Costume Or Enhancement Selection Specification 0200 will often be sent only once, however, it is possible to change this specification in the middle of a Fantasy Video show or conversation. The next essential information is the Costume Configuration Vector 0210, consisting of joint and actuator values, and parameters specifying poses and other configurations. The Presentation System 0012 uses this information to help determine the pose of the costume for constructing the presentation 0050. Other required information includes the Wiring Information 0215, which is typically a table that determines what changes should be made in the Costume Model 0205 based on changes in the Costume Configuration Vector 0210. Although unusual effects may be achieved by modifying the Costume Model's 0205 size or color based on sound or positional information, typically this will be restricted to corresponding joint and configuration values. In Figure 3A, these are illustrated using the Left Elbow Bend Joint 0291, the Left Shoulder Rotation Joint 0292, the Left Hip Rotation Joint 0293, and the Left Knee Bend Joint 0294. A listing of the joints used by Poser 3.0™, a current-art figure animation system that runs on a single computer and takes commands directly from an editing viewer, is presented in Appendix 1.

It is also possible for the Imaging System 0010 to send a 3D model of the actual user, in a static standardized pose or in the current pose, to the Presentation System 0012.

Other essential information for this paradigm includes a Camera Information 0220 packet specifying the locations and characteristics of any virtual cameras used. Location, consisting of position plus orientation, can be specified by using such technology as 4x4 homogeneous transformation matrices, Euler angles, or quaternions. Camera characteristics include lens angle, aperture, focal length, and virtual filters. Codes can also be used to switch back and forth between previously-specified virtual cameras.

The resulting Camera Information 0220 is used to select and possibly configure one or more virtual Presentation Cameras 0225 for use in computer graphics.

Similarly, a Lighting Information 0230 packet specifying the locations of characteristics of any virtual lights used may be included as part of the essential information. Lighting characteristics include such things as color distribution, fall-off, barn-door angle, shape, extent, focus or diffusion, etc. Codes can also be used to switch back and forth between previously-specified virtual lights. The resulting Lighting Information 0230 may be used to select and possibly configure zero or more virtual Presentation Lightings 0235 for use in computer graphics.

A Texture Information 0240 packet specifying the locations and characteristics of any virtual textures used may also be included as part of the essential information. Texture characteristics include such things as 2D color distribution on a patch, bump-map distribution, reflectance distribution, transparency distribution, and offset distribution, etc. Codes may also be used to switch back and forth between previously-specified virtual textures. The resulting Texture Information 0240 may be used to select and possibly configure zero or more virtual Presentation Textures 0245 for use in computer graphics.

A Literal Texture Information 0250 packet specifying characteristics of literal textures and images may also be included as part of the essential information. This includes such things as literal images of the eyes, mouth, or face of the current user(s) 0001. The result may be communicated through the distribution channel 0011 to form zero or more Presentation Literal Textures 0255.

In this way, essential information is sent over the distribution channel 0011. On the viewer's side, the presentation construction subsystem 0012 uses this information to construct a presentation 0050 standing for the user. If merely a slight change or augmentation is being made, the subsystem uses a model of the actual user to construct the presentation. If a replacement is being made, as is shown here, the subsystem uses a virtual costume for presentation construction. The virtual costume consists of a costume model 0205 in a standard pose; a set of joint or actuator variables; and "wiring information 0215", being instructions as to how to modify the appearance of the model based on values in the set of joint or actuator variables. The presentation construction system 0030 takes the virtual costume model 0205, takes the set of joint and actuator values (the Costume Configuration Vector 0210) from the transmitted information, uses the wiring information 0215 to modify the appearance of the model, changes the lighting, camera, and texture models as specified and uses the changed values to create a presentation of the user, performs at the same time a similar process on the environment, and thus creates a presentation of the user plus environment.

A similar system is used for auditory, force, and other sensor modality features. For instance, the user's voice is tracked; features in the voice are abstracted into a model of what is going on in the scene, which is abstracted into essential information; the information is transmitted over the distribution channel;

a virtual costume is used to construct an image standing for the user; and the image is presented to the viewer.

More than one user can be composited into the same presentation. The multiple users can come from the same imaging system, or they can come from multiple imaging systems in a multiplex arrangement.

Fig. 3B shows another embodiment of the methods used for imaging and presentation. These methods are based on perspective view morphing, which supports morphing between two or more 2D images when the images have been overlaid with correspondence points, and in which the resulting 2D image appears as it would if the morph had been performed in 3-D with two 3D objects of different orientations. This capability supports a Fantasy Video circuit. In the drawing, the explanation is restricted to face presentation for simplicity. The image information representation subsystem 0021 uses a library of trained eigenfaces that attempt to span the space of all significant facial appearances for the user. An eigenface is a characteristic face that marks one axis of the space. Each eigenface has a stored representation, composed of a 2D image plus an overlay of feature locations in the image. The Image Acquisition Subsystem 0020 acquires an image of the actual user 0001. The input user appearance, called the User's Current Image In User Face Space 0301, is abstracted by finding which eigenfaces it is closest to, and measuring its morph coordinates known as Face Space Coordinates 0390. For instance, in the drawing the User's Current Image In User Face Space 0301 is closest to stored eigenfaces A, B, C, and D, being Point A in User Face Space 0311, Point B in User Face Space 0312, Point C in User Face Space 0313, and Point D in User Face Space 0314. Then the corresponding morph Face Space Coordinates 0390 are measured to be 0.2, 0.4, 0.8, and 0.6, respectively. The morph Face Space Coordinates 0390 are proportional distances in face space from each of the nearest eigenfaces, such that if a perspective view morph were performed between these nearest eigenfaces using the morph coordinates, the input appearance would appear. Morph coordinates may be obtained by convolving the actual image over the face-space of potential morphed images, and selecting the best match. The obtained morph Face Space Coordinates 0390, along with their coordinate axes being codes for the nearest eigenfaces, are thus declared to be an abstracted eigenvector that contains the essential information for representing the image.

This information is sent to the presentation construction subsystem 0030. The presentation construction subsystem 0030 could augment or modify the appearance, but here we show a replacement with a dragon face. The presentation construction subsystem 0030 has access to a set of costume eigenfaces, representing a space of corresponding poses. The abstracted eigenvector being the Face Space Coordinates 0390 is used to create a presentation of the user 0050. First, the identities of the nearest eigenfaces, being A, B, C, and D, are determined from the information. These are Point A in Presentation Face Space 0341, Point B in Presentation Face Space 0342, Point C in Presentation Face Space 0343, and Point D in Presentation Face Space 0344. These eigenfaces are retrieved and used to conceptually

construct the local axes of a face space. Then the eigenvector Face Space Coordinates 0390 are used to determine a point in this face space. At this point, a perspective view morph is performed between the selected eigenfaces, using proportion values determined by the eigenvector. This results in a 2D image of the dragon's face, called the Derived Current Image in Presentation Face Space 0350, oriented such that it appears as if it were rotated proportionally in 3D between 3D representations corresponding to the actual 2D eigenface images. This image is declared the Presentation of the User 0050 and is used for output presentation on a 2D display device. The operation can be performed twice to generate two stereo-3D images. A similar operation can be performed for phonemes or other sound features, or for other sensory modalities supporting morphing.

Figure 3C illustrates a typical paradigm for essential information and presentations based on phonemes for sound images. The user(s) 0001 is speaking. The Imaging System 0010 uses a microphone for an image acquisition device 0023 and a corresponding Means for Digitizing Images 0024 as shown in Fig 1F to gather a sound image of the user. Then the imaging system 0010 uses its image information representation subsystem 0021 to run a phoneme-recognition algorithm to acquire and represent features of the user's sound image. These features are declared to be the essential information for the sound image. This essential information is sent over the distribution channel 0011 to the Presentation System 0012 on the other side, where it is used to construct a presentation of the user 0050.

In this paradigm, the essential information consists of all of the information needed to help create a sound image of the user, including phoneme identity, pitch, loudness, and duration. Phoneme Information Packets 0370 are used to represent this essential information in the Imaging System 0010. Then this essential information is shipped to the Presentation System 0012 and used along with the Voice Font 0380 formatting information to create the Presentation Phonemes 0375. These are then declared to be the Presentation of the user 0050. Thus the user can communicate while changing the sound image.

Fig. 4 demonstrates example classes of types of changes that can be performed on the appearance of any users and/or the environment(s). The user 0001 is shown on the left as input in each case; on the right is the presentation of the user 0050 resulting from a change or enhancement of a particular type. The Presentation of the User using Replacement 0451 demonstrates "Replacement", which results in the appearance being replaced by another appearance. Various types of replacement are possible, such as head replacement, face replacement, body replacement, hand replacement, etc. Here we see the user's appearance being replaced by a virtual costume of a dragon. The Presentation of the User using Augmentation 0452 demonstrates "Augmentation", which results in the appearance having extra features added in to it. Here we see the appearance of the user being augmented by adding a halo, a beard, and a set of wings. Different types of augmentations and other changes are discussed elsewhere in this work. The Presentation of the User using Modification 0453 demonstrates "Modification", which results in existing features being changed. Here we see the presentation of the user's legs and head being modified by

elongating them, while the torso is compressed. The Presentation of the User using Deletion 0454 demonstrates "Deletion", which results in removing existing features. In this illustration, the user's hair and legs have been cleaned up and deleted from the presentation. The Presentation of the User using Filtering 0455 demonstrates "Filtering", which is similar to modification, but tends to be applied in a nonsemantic manner; that is, it uses less intelligence when making its changes. ("Turn everything blue" rather than "make the legs robotic"). Filtering in visual images tends to work with global changes of the 2D appearance, rather than local changes of the 3D appearance. Many appearance filters for static images are already being sold with Adobe(tm) Photoshop(tm). Here we see a swirl filter being applied to the appearance of the user, followed by a mosaic filter. The Presentation of the User using Overriding 0456 demonstrates "Overriding", in which the physical appearance of the user stays the same, but the movements and actions are changed by a filter, a program, or a third-party controlling agent. Here the left arm and leg have been overridden to show the presentation of the user using overriding 0456 holding his left arm down and bending his left knee. The Presentation of the User using Restaging 0457 demonstrates "Restaging", in which the camera and lighting parameters are changed. "Restaging" also includes "Repositioning", in which the location of the camera optionally stays the same, but the global location of the user in the presentation is changed. Here we see the lighting being made more stark to cast a shadow behind the presentation of the user using restaging 0457, the depth of focus being tightened so that the arms are out of focus, a mirroring transform being applied to the camera, and the location of the user being repositioned slightly to the left. The Presentation of the User using a Combination of Techniques 0458 demonstrates that the enhancement changes can be performed in combinations as well. Here we see the head being replaced by a dragon costume; the body being augmented with a beard, halo, and wings; the legs being modified by lengthening, along with a shortened trunk; and the lighting and camera restaged to cast sharper shadows and point the presentation to the left.

Changes can be performed on the 3D model, 2D image, lighting model, camera model, texture model, sound model, voice font, action variables, or any other information used to construct the presentation. It is also possible to base changes on cross modalities. For instance, the location and color of objects can be modified based on the pitch of the sound, which results in objects that dance to music. Or, the pose of the jaw and lips can be driven based on perceived phonemes, which results in lip-synch mouth-tracking in the presentation that does not require geometric pose information to be transmitted across the distribution channel.

Fig. 5 shows some of the features typically used to encode the essential information representing the facial configuration of the user. A typical information packet would include: the Top of Head Y Coordinate 0510; the Center of Head X Coordinate 0520; the Outside Corners of Eyes X and Y Coordinates 0531; the Inside Corners of Eyes X and Y Coordinates 0532; the Iris Center Gaze Angle 0540; the Corners of the Mouth X and Y Coordinates 0550; the Top of the Top Lips 0560; the Bottom of the

Bottom Lips 0570; and the Bottom of the Chin Y Coordinates 0580. Coordinates may be given measured from image-centric or world-centric origins.

Fig. 6 presents example distribution channels 0011. A distribution channel 0011 sends the essential information from one or more imaging systems 0010 to one or more presentation systems 0012. The first example is a Local Area Network or Wide Area Network Distribution Channel 0011a, along with the interface network cards necessary to send the information across the network. An imaging system sends the information to its network card 0610a, which sends the information over the network 0605 to the remote network card 0610b, which then gives the information to the presentation system. Next there is a telephone network distribution channel 0011b. This distribution channel requires a modem 0620a, which uses a telephone line 0616a to interface into the telephone cloud 0615; the details of the inside of the telephone network cloud 0615 are not required to be known. A similar telephone line 0616b interface accepts the information for a modem 0620b for the presentation system. Modems are often used in computer-to-computer communication. Next there is an Internet connection distribution channel 0011c. The imaging system uses a modem 0620a to talk with an Internet Service Provider (ISP) 0626a, usually over the local phone network. This ISP 0626a goes through the Internet 0625 to a remote ISP 0626b, which then typically connects with the remote presentation system using some kind of modem 0620b again. This embodiment makes it easy to create a repeater or central server that takes incoming essential information from one or more users on one or more presentation systems, puts the information together into one or a few streams (also called "channels", in the sense of a TV channel or TV network that can be selected by the viewer), and sends it out to one or more viewers on one or more presentation systems. Next there is computer tape, diskette, or removable hard disk as a distribution channel 0011d. An imaging system uses a means for copying media 0631 to write its essential information to computer tape, diskette, or removable hard disk 0635. The computer tape, diskette, or removable hard disk 0635 is possibly duplicated or replicated at a factory, representing another means for copying media 0631, and then it is mailed out or carried to one or more viewers via mail or personal transport 0636. A viewer loads the computer tape, diskette, or removable hard disk 0635 into the media reader 0632 of the local presentation system, where the information is pulled off the computer tape, diskette, or removable hard disk 0635 into the system. This distribution channel differs from the previous ones in that it is not inherently bi-directional, and that it is non-real-time. After this, there is the cellular telephone network distribution channel 0011e. A cellular telephone uses an internal modem 0620a to modulate the essential information into radio waves at one of the cellular telephone frequencies, which are then picked up by the cell phone network 0645 and sent across the telephone network to the presentation system's side. If the presentation system is also on a cell phone, it uses a receiving modem 0620b to receive the information; or it can accept the information from the telephone network by using a more conventional phone line 0616b and modem 0620b. One useful embodiment is to build the cell phone and imaging system, or cell phone and presenting system, as a unit device. Then the imaging system 0010 or presentation system 0012 can use the sending and receiving

capabilities of the cell phone circuitry directly. It is also possible to build a device that has the capabilities of a cell phone, an imaging system, and a presenting system. This can be used as a Fantasy VideoCellPhone for two-way communication.

Next, there is a CD-ROM, CDR, or DVD-based disc distribution channel 0011f. This is similar to the computer tape, diskette, or removable hard disk as a distribution channel 0011d, in that it is non-real-time. The imaging system 0010 sends its information to a manufacturing plant 0641, which duplicates a disc 0655. The disc 0655 is distributed in stores or by mail, 0656. The viewer 0005 uses an appropriate disc reader 0642 to read the disc 0655 and obtain the information for the presentation system 0012. Next there is the Internet Telephone distribution channel 0011g. This uses the Internet to carry telephone signals; the telephone signals can then be broken out of the Internet at one end or the other, and sent across regular telephone lines. A modem 0620a can be used to interface in to the Internet Telephone Network 0665, and another modem 0620b is used at the other end. Next, there is a cable TV network used as a distribution channel 0011h. The information is sent to a cable broadcasting station 0651, which sends it down the cable 0675 to a cable receiving module 0652. Cable is useful for broadcasting to many receiving presentation systems, but it is not very useful for information flow in the reverse direction. Next there is radio waves in the television or radio spectrums as a distribution channel 0011i. Here the information is encoded in a television or radio signal 0685 which is broadcast using a broadcasting station 0661, and picked up by a television or radio receiver 0662. It is possible in the case of television to broadcast the Fantasy Video information inside the retrace signal, and thus be able to carry a regular television broadcast on the same signal at the same time. The final example distribution channel is files carried on a Web page. This channel is one-to-many; in fact, it is easy for multiple presentation systems to read information from one source at the same time. In a non-real-time implementation, essential information from an imaging system is sent to a file, which is then placed on a Web page server 0671. The contents of the Web page are available over the Internet 0625. A viewer 0005 uses a presentation system 0012 module built in to a Web browser to call up the page and download the file, by means of a Web client 0672, and view the presentation as it streams over the Internet 0625. A real-time implementation uses a server 0671 to broadcast the information on the fly, as it comes in from the imaging system 0010; the presentation-system browser plug-in with Web client 0672 is essentially the same.

Fantasy Videophone conversations, Email or Fantasy Video TV/movies may be sent over any of these distribution channels. For example, a cellular telephone network distribution channel 0011e could be used to carry Fantasy Video Email.

Fig. 7 shows various example topologies for multiple-user connections of Fantasy Video Circuits. A circuit is alternately one-to-one, one-to-many, many-to-one, or many-to-many, between originating imaging systems 0010 and receiving presentation systems 0012. For instance, a typical single-actor Fantasy Video TV broadcasting arrangement probably uses a one-to-many multicast Broadcast Topology

0740 circuit between one imaging system 0010 and many presentation systems 0012. Alternatively, this same Broadcast Topology 0740 could be implemented as a set of point-to-point connections between the imaging system and each presentation system, as there are no restrictions against having multiple distribution channels simultaneously emanating from the same imaging system. In this discussion, it will be assumed for the sake of convenience that each imaging system serves only one user. However, it is quite possible for a single imaging system to have more than one user, or even no users at all. Similarly, it is possible for a presentation system to have more than one viewer, or no viewers at all.

Each topology actually has a few versions, one in which the circuits are two-way, and one in which the circuits are one-way in a particular direction. Mixes are also possible. For the two-way circuits, a particular station will consist of both a Fantasy Video Sender 0008 and a Fantasy Video Receiver 0009 put together into a Fantasy Videophone Station 0007. It is also possible to have a local presentation system 0012 feeding back the results of an outgoing imaging system 0010 essential-information stream, without having a two-way circuit where information is also coming in off of a distribution channel 0011.

The first example topology is the Ring Topology 0710, in which each station is connect to its two neighbors. Each station must forward the presentation information on to the next. Singly-linked or doubly-linked rings are possible; a doubly-linked ring can require half the circumference count in delay time to relay a signal, whereas the singly-linked ring can require the entire circumference count in delay time if a viewer is directly behind a user in the ring. Rings are relatively slow, but relatively easy to implement, as only one or two connections are needed for each station.

Next there is a Mesh Topology 0720. Each imaging system 0010 or presentation system 0012 is connected directly to every other one in the group. This results in the shortest delay, but the most connections to support. The Star Topology 0730, on the other hand, uses a central repeater or server to relay signals from one system to all the others. This topology is popular for cyberspaces, where all of the users' avatars in a particular scene are assembled by the central repeater into one signal and sent as a whole to the viewers. Often the central server will be run by a commercial third-party company.

In the Star Topology 0730, the central repeater typically does not contain a user's imaging system 0010. The "client/server" or Broadcast Topology 0740, on the other hand, although apparently similar, typically will contain an imaging system 0010 on the central server and typically will not act as a repeater between the satellite stations. The Broadcast Topology 0740 can be one-to-many, in which case it is useful for Fantasy Video TV--a Fantasy Video movie signal is sent out to many viewers, perhaps starting at different times if it is a pay-per-view system. Alternatively, the system can be many-to-one, if the system is used for marketing purposes where multiple users 0001 report in to a single viewer 0005. Finally, the system can be one-to-many two-way, for instance for a taxicab dispatching application where there is a central dispatcher that many taxis each talk with.

The Multi-To-Many Relay Topology 0750 is a special case of the Star Topology 0730 where a few users 0001 on the left are connected through a central relay station to many viewers 0005 on the right of the drawing. The connections can be one-way or two-way; however, typically the viewers 0005 on the right cannot view each other, but only the users 0001 on the left. This topology is useful for sports arenas, plays, and classroom dramas. An alternative embodiment is the Multi-To-Many Direct Topology 0760, where again each user 0001 on the left is connected with each viewer 0005 on the right, and perhaps with each other, but the viewers 0005 on the right do not get to view each other.

Finally, we have the Network Topology 0770, having switches, routers, and repeaters to implement the connections in an extremely general fashion.

These examples illustrate a few of the many possible connection topologies that are possible to use in the distribution channel 0011 embodiment when one or more users 0001 using one or more imaging systems 0010 communicate in a one-way or two-way fashion with one or more viewers 0005 using one or more presentation systems 0012.

Figs. 8A, 8B, 8C show a few of the many possible presentation devices 0040 that can be used in a Fantasy Video Circuit. Presentation devices 0040 can support 2D or 3D presentations; they can make presentations on screens, on paper, or carved into solid materials; or, in the case of the laser eye projector, they can make a presentation directly into the eye without a physical manifestation at all. Not shown is a speaker or headphones set for audio presentations, and an active force-driven joystick for haptic feedback presentations.

In addition to the 3D head-mounted display, which uses direct projection on two tiny screens close to the eyes, it is popular to use 3D stereo glasses having LCD light valves in them to view a computer monitor that shifts back and forth between the left eye and right eye view at 60 Hz.

Of special note are the flat-panel high-definition TV monitors or screens 0040d, which can be mounted on a wall, and the handheld cellular 0040n or wristwatch 0040p videophones. As is discussed, both of these classes of devices require re-placement of the virtual camera position, because the actual camera position is either uncomfortably off-axis or too close or both when imaging a typical Fantasy Videophone user.

Any one or more of these presentation devices 0040 can be used in a Fantasy Video Circuit, Fantasy Video Receiver 0009, or Fantasy Videophone Station 0007, among other embodiments presented in this work. There are two main classes of applications: one-way receiving applications in which the viewer(s) 0005 can watch a "TV" or "movie", which are based on the Fantasy Video Receiver 0009 and in which an Image Acquisition Device 0023 is optional; and, two-way communication applications in which the viewer(s) 0005 are also user(s) 0001, which are based on the Fantasy Videophone Station 0007 and in which an Image Acquisition Device 0023 is mandatory. The artist has included a camera in the depiction

of objects in which it would typically be found, but keep in mind that any of the presentations devices 0040 can be used in configurations with an Image Acquisition Device 0023 or without one. Only the wrist-watch TV 0040o1 and the wrist-watch videophone 0040p are explicitly shown separately, because of their importance.

The intent is not to limit the presentation device 0040 to the devices shown here, but to incorporate any present and future devices that support a means for displaying a sensory presentation.

Presentation devices 0040 illustrated in Figs. 8A, 8B, and 8C include: a computer monitor 0040a; a television set or television monitor 0040b; a high-definition television 0040c; a flat-panel display 0040d, such as is mounted on a wall; a 3-D head-mounted display 0040e; a system comprising a 3-D movie or 3-D computer monitor display 0040f, using lenticular lens gratings or LCD light-shutter devices in a flat panel or in viewers' glasses; a hologram-making device 0040g; a building-sized display sign 0040h; a billboard 0040i; a printer, color printer, photo printer, hologram film printer, hologram foil stamper, or color separation negative printer 0040j; a picture-phone, screen phone, or videophone, including desktop phone 0040k1 and pay-phone 0040k2 styles; a TV set-top device connected to a TV set or monitor 0040l, including cable boxes and family game computer systems; a fax machine 0040m; a cellular TV, cellular picture-phone or cellular videophone 0040n; a wrist-watch TV 0040o1 or portable TV 0040o2; a wrist-watch picture-phone or videophone 0040p; a laser-driven sculpting device 0040q1 or N.C. router-based sculpting device 0040c, yielding output in wax, plastic, wood, metal, ice, or steel; an LCD, dye, or plasma screen 0040r; direct-to-magazine printers 0040s; a laser-based device that projects an image directly onto the viewer's fovea from glasses or a head-mounted device 0040t1, or laser-based device that projects an image directly onto the viewer's fovea from a desktop 0040t2; a headset or wearable computer or fabric computer 0040u; a window display on a vehicle such as an automobile, truck, bus, plane, helicopter, boat, tank, motorcycle, crane, etc. 0040v; a neural transmitter that creates sensations directly in a viewer's body 0040w; a computer-based movie projector or projection TV 0040x; a hand-held game device 0040y; a palmtop, laptop, notebook, or personal assistant computer 0040z; a screen display built into a seat or wall for use in the home, on airlines, inside cars, or in other vehicles 0040aa; a computer monitor used in an arcade game or home computer game 0040bb; and, a screen or speaker integrated with an appliance such as a refrigerator, toaster, pantry, or home-control system 0040cc.

The various Figs. 9A, 9B, 9C, 9D illustrate alternative placements in a Fantasy Video Circuit for an optional auxiliary library of enhancement changes, environments and virtual costumes. This Library of Formatting Information 0900 supplies formatting information that is used by the Presentation Construction Subsystem 0030 of a Presentation System 0012 to help make a Presentation 0060 for a viewer 0005. When a selectable Library of Formatting Information 0900 is used, it is useful to optionally introduce a Library User-Interface for the User 0910 and/or a Library User-Interface for the Viewer 0920, which are used to make selections from the Library 0900 and to negotiate as to which formatting information is actually used.

Both the Library User-Interface for the User 0910 and the Library User-Interface for the Viewer 0920 must have some sort of communication path established to a Library of Formatting Information 0900; this path is used for sending commands and obtaining listings, icons, and appearances. The usage is straightforward; instead of a single enhancement filter or virtual costume, a Library 0900 allows a user 0001 and/or a viewer 0005 to select the enhancements of their choice. A Library User-Interface for the User 0910 or for the Viewer 0920 should use its communication path to query the Library 0900 as to which enhancements, including changes, avatar costumes, environments, etc., it can provide. These should then be presented in a list or menu to the user 0001 or viewer 0005 respectively. The Library User-Interface for the User 0910 or for the Viewer 0920 should then accept menu requests as to which enhancements are desired, and send these requests to the Library 0900. The Library of Formatting Information 0900 then makes the requested formatting information available to the viewer's 0005 Presentation Construction Subsystem 0030 in a direct or remote manner, where it is used to help construct the Presentation(s) 0060 consisting one or more Presentation(s) of the User(s) 0050 and Presentation(s) of the Environment of the User(s) 0052. Libraries 0900 connected directly to or part of a viewer's 0005 presentation system 0012 can contribute formatting information in a direct manner, whereas Libraries 0900 connected to or part of a distribution channel 0011 or an imaging system 0010 must communicate their formatting information to the viewer's 0005 Presentation Construction Subsystem 0030 in an indirect manner, e.g. by going through the distribution channel 0011. It is not necessary to have only one Library of Formatting Information 0900, there can be multiple Libraries 0900 at any, most, or all of the major points in the system (circuit).

Fig. 9A shows a Library of Formatting Information 0900 attached to the imaging system 0010. The user 0001 selects a costume, environment, enhancement method, or other formatting information directly from the Library 0900, by means of the Library User-Interface for the User 0910, which formatting information is then sent along with or ahead of the actual conversation as information for creating a presentation 0060. If the viewer 0005 wants to override the selection with the viewer's 0005 choice, the viewer 0005 has to send a request back over the distribution channel 0011 or other communication means to the Library 0900 in order to fetch the new environment or costume. This is done using the Library User-Interface for the Viewer 0920. Thus, there is a negotiation that goes on, between the user's 0001 choice of formatting information, and the viewer's 0005 choice of formatting information. Various priority or lock-out schemes are possible, but the easiest and simplest scheme is to merely allow either the user 0001 or the viewer 0005 to be able to change the enhancements used in constructing the presentation 0060 at will. If the user 0001 does not desire that the user's 0001 actual appearance be presented, then a more advanced scheme could allow the viewer 0005 to request any formatting information that the viewer 0005 desires, as long as it includes a replacement change for the presentation of the user's 0050 face.

Fig. 9B shows the Library of Formatting Information 0900 on the viewer's 0005 side. The user 0001 has to request a menu from the Library 0900 or has to have already cached a menu of the possibilities, in order to select a request for a favorite costume. Then the user 0001 causes instructions to be sent to the

Library 0900 by means of the Library User-Interface for the User 0910 to load that selection in the presentation system 0012. In the case of the viewer 0005 overriding the selection, this is done in a more straightforward manner, as the viewer 0005 has merely to load the new overriding selection from the Library 0900 directly. This is accomplished by means of the Library User-Interface for the Viewer 0920.

Fig. 9C shows the Library of Formatting Information 0900 as part of a third-party support system that is actually part of the distribution channel 0011, as in the case of a cyberspace company. Both the user 0001 and an overriding viewer 0005 have to send a query through the distribution channel 0011 in order to reach the Library 0900; the Library 0900 then sends its presentation construction information through a distribution channel 0011 to be loaded into the presentation system 0012.

These alternatives are not necessarily mutually exclusive. For instance, it is possible to get a costume for the user's body from a Library 0900a on the user's 0001 side, a costume for the user's head and face from a Library 0900c on the viewer's 0005 side, and a new environment from a Library 0900b on a third-party support system in the distribution channel 0011, all for the same presentation 0060, as illustrated in Fig. 9D.

Fig. 10 shows an embodiment that is a Fantasy Video TV Broadcasting Station with Multitrack Editing. The system is designed to create and broadcast a Fantasy Video Movie which is basically the same as a Fantasy Video Email message, that is, it is typically a file that contains a stream that was produced previously off-line. The user(s) 0001 who produce movement information for the Fantasy Video Movie are called "actors". A Fantasy Video TV Broadcasting Station can have one actor 0001, or it can have multiple actors 0001. Typically each actor 0001 will get his or her own imaging system 0010.

Instead of sending the essential-information stream directly out onto the distribution channel 0011, the Broadcasting Station records it locally, using a Recording System 1010 coupled with a Storage System 1020. The Recording System 1010 accepts the stream of essential information from the imaging system 0010 and records this stream using the Storage System 1020, which will typically be a hard disk or magnetic tape. These Systems will typically be embodied in the same device. One or more viewers acting as "editor" or "director" 0006 then edit various streams of information using a multitrack editor and layer the streams down into a single stream. The multitrack editor has an Editing System 1030, an optional Editing System User Interface 1040, and an optional Presentation System 0012 for viewing the results of the edits locally. The Editing System 1030 accesses streams of information from the Storage System 1020, displays them on the editor/director's presentation system 0012, accepts editing commands from the Editing System User Interface 1040, and creates an edited stream of information, composed of new essential information, formatting information, and commands, etc., which is then typically saved in a Storage System 1020. These results are finally made available for broadcasting over the distribution channel 0011. The information is then sent to the Presentation System 0012 as usual, where it is finally presented to a viewer 0005. Fantasy Video TV typically uses one-way circuits, so that the viewer 0005 can

watch the presentation 0060 but cannot converse with the user(s)/actor(s) 0001. In the case of a talk show or certain other types of shows this will not always be the case, however.

Fig. 10B shows an embodiment of a Fantasy Video Recorder. This is a basic component in the Fantasy Video Broadcasting Station and the Fantasy Video Email Sender. The Presentation System 0010 abstracts a stream of information describing one or more users 0001 as usual. However, the information stream may not be sent directly out to the distribution channel, but rather is routed to a Recording System or Means for Recording 1010. This records the stream into a Storage System 1020. The Means for Recording 1010 is typically a simple buffer that supports writing out onto a file; the Storage System 1020 is typically a hard disk or a reserved buffer in memory. The results of the Fantasy Video Recorder are a saved stream that can be edited later.

Fig. 10C shows an embodiment of a Fantasy Video Editor. This is a basic component for working with previously-recorded streams of essential information. An Editing System 1030 drives a Presentation System 0012 that shows replays of the tracks being edited. An Editing System User Interface 1040 controls the Editing System 1030. The Editing System 1030 takes tracks and streams of information to edit out of a Storage System 1020, and writes finished resulting streams of information back in to the Storage System 1020. The results of the Fantasy Video Editor are saved streams of information that have been edited.

Fig. 11 shows an embodiment that is a Fantasy Video Email system. Instead of sending the essential-information stream out directly, it is diverted and recorded into a buffer, by means of a Recording System 1010 and an optional Storage System 1020. The Storage System 1020 buffer could be in a computer's memory, or it could be in a more permanent medium such as a hard disk or magnetic tape. The buffer is optionally made available for playback and re-recording, or perhaps even editing in advanced email systems. This is done by optionally including an Editing System 1030 combined with an Editing-System Presentation System 0012e and an Editing-System User Interface 1040 to form a Fantasy Video Editor. The user 0001 optionally plays back the recorded e-mail message by using the Editing-System User Interface 1040 to command the Editing System 1030 take the e-mail message out of the buffer in the Storage System 1020 and send it to the Editing-System Presentation System 0012e for presentation to the user 0001 for verification. The user 0001 can then modify and edit the Fantasy Video e-mail message using the Editing-System User Interface 1040 and the Editing System 1030, which writes the edited message back into the Storage System 1020. The user 0001 can also use the Editing-System User Interface 1040 to request the Editing System 1030 to direct the Recording System 1010 to re-record portions or all of the message by again using the Imaging System 0010 and the Storage System 1020. When the user 0001 is satisfied with the message, the message is sent over the distribution channel 0011. Although email traditionally uses the Internet, there is no reason why the email should be restricted to using that distribution channel 0011. The message is sent by using a Fantasy E-mail Channel Sending Subsystem

0015e which spools the message from the Storage System 1020 to the distribution channel 0011. The distribution channel carries it to the viewer's 0005 side, where it is typically picked up by a Fantasy E-mail Channel Receiving Subsystem 0018e and spooled or copied into an Incoming E-mail Message Storage System 1020m. (The Fantasy E-mail message can be sent from the distribution channel 0011 directly to the viewer's 0005 presentation system 0012 if the viewer 0005 is attending and does not wish a record of the e-mail message.) Assuming the e-mail is recorded or buffered in the Incoming E-mail Message Storage System 1020m, the viewer 0005 can then select which e-mail message to view by means of the Email Playback User Interface for the Viewer 1060, which informs a Message Playback System 1050 to take the selected message from the Incoming E-mail Message Storage System 1020m and play it on the viewer's presentation system 0012. In a two-way system, the viewer can then reply back. Fantasy Video Email has the advantage that it will typically be much smaller than corresponding normal video email, due to the inherent high compression rate. Note that either visual images or aural images or both may be sent, stored, and presented.

Fig. 12 shows an embodiment that is a Fantasy Videophone Answering Machine. The Answering Machine must run on a two-way circuit. When a call comes in from a caller, the Answering Machine Control Component 1200 attempts to signal the user by "ringing". This can consist of physical ringing, a screen flashing, a vibrator buzzing, or some other kind of signal that lets the user know a call is coming in. If the user-acting-as-viewer 0003 does not "pick up" within a certain amount of time, such as 4 rings, the Answering Machine Control Component 1200 activates the answering machine features. First, an optional prerecorded message is played for the caller, by means of an Outgoing Message Playback System 1050o that relays the outgoing Fantasy Videophone message from the Outgoing Answering-Machine Message Storage System 1020o to the outgoing distribution channel 0011, where it can be seen by a calling viewer 0005. This message was previously recorded by the user-acting-as-viewer 0003 by using the Outgoing Message Recording System 1010o that takes its input from the Imaging System 0010 and records the outgoing message into the Outgoing Answering-Machine Message Storage System 1020o. In an advanced system, the Answering Machine Control Component 1200 has a means of identifying the caller, e.g. caller ID, so that it can choose an appropriate one of several possible outgoing answering messages to play for the caller, based on the caller's identity. After the Answering Machine Control Component 1200 finishes playing the outgoing message, it activates an Incoming Message Recording System 1010i that records the Fantasy Videophone stream coming in from the distribution channel 0011 to the Incoming Answering-Machine Message Storage System 1020i, which stores the incoming message for later playback. Later playback is accomplished by the called-person user-acting-as-viewer 0003 using an Incoming Message Playback User-Interface For The Viewer 1060i to select a desired incoming message and control an Incoming Message Playback System 1050i. The Incoming Message Playback System 1050i takes the selected message from the Incoming Answering-Machine Message Storage System 1020i, and sends it to the called-person user-acting-as-viewer's 0003 presentation system 0012 for presentation viewing.

The Answering Machine is quite similar to the Email configuration, except that the recording of the incoming messages is done on the remote side from the caller instead of on the local side.

Fig. 13 shows a Formatting Information Editor that is used by a viewer acting as "editor" or "director" 0006 to interactively construct a creative virtual costume model 0205, appearance change or enhancement for a user 0001, or appearance change or enhancement for an environment 0002. The user 0001 will typically be the same as the viewer-acting-as-"editor"-or-"director" 0006 controlling the Formatting Information Editor. The viewer-acting-as-"editor"-or-"director" 0006 positions a user in front of an optional imaging system 0010, or otherwise gets a sample information stream of essential information from a Storage System 1020. This could be as simple as a single pose of a person standing still. The viewer-acting-as-"editor"-or-"director" 0006 then views the sample information on a presentation system 0012, while editing and refining the formatting information being used by the presentation system, by means of a Formatting Information Editing System 1031 controlled by a Formatting Information Editing System User Interface 1041. The Formatting Information Editing System 1031 typically works with a local storage buffer in a Storage System 1020 for maintaining the formatting information while it is being created and worked on. When the user is finished adjusting the formatting information, it is saved into a Library of Formatting Information 0900 for future use.

Fig. 14 illustrates one of the many advantages of the invention. Previous art, which transmits literal video images, must deal with actual appearances as they are acquired, in an unenhanced fashion. The top portion of Fig. 14 shows a single user 0001 using either a wrist-watch videophone with a built-in camera Image Acquisition Device Mounted In A Wrist-Watch Videophone 0023w, or using a videophone having a camera Image Acquisition Device Mounted Above A Screen In A Videophone 0023c positioned above a television monitor 0040b. The unenhanced images that these yield are ugly, due to bad camera locations and parameters. A wrist-watch videophone requires a wide-angle lens in order to acquire the entire face of the user 0001, but this results in fish-eye distortion; whereas, a camera mounted above a screen must necessarily be looking down on the user 0001 if the user is to maintain natural eye-contact with the presentation being watched on the screen 0040b. The results are shown below, as the Original Sensory Image of a User from an Up-Close Wrist-Watch Camera 0265, and the Original Sensory Image of a User from a Camera Above A Screen 0266. Again, the prior art is forced to transfer these literal images. This invention can overcome these problems by using a "restaging" enhancement that shifts the position of the virtual camera and modifies its parameters, while optionally leaving the other appearance parameters of the rest of the scene undisturbed. For instance, for the wrist-watch videophone, the virtual camera location can be shifted back three feet and given a regular lens; while for the camera above the screen videophone its virtual camera location can be shifted downwards by a foot and a half and then rotated to track the user 0001. The results of these presentation enhancements are shown in the Presentation of the User with Perspective Corrected 0050z, which is then enjoyed by the viewer 0005. Note that it is not necessary to

introduce fantastical replacements in the presentation 0050; simply enhancing the presentation 0050 by restaging but leaving the rest of the appearance alone is claimed as part of the general enhancements.

Figure 15A illustrates one embodiment of a Fantasy Videophone Station communicating with another Fantasy Videophone Station to form a two-way Fantasy Video Circuit. A user also acting as a viewer 0003 sits in front of an Imaging System 0010 having hardware of an Image Acquisition Device mounted above the Screen 0023c and a computer that is family game play station acting as a "TV set-top device connected to a TV set or monitor" 0040l, connected to a television monitor 0040b. The monitor displays a presentation of the remote user 0060. The "TV set-top device connected to a TV set or monitor" 0040l runs the software of the Image Information Representation Subsystem 0021 and the Presentation Construction Subsystem 0030; it is connected to a Internet Cloud or POTS ordinary Telephone Network Cloud 1500 by means of a Telephone Wall Socket 1510. A similar Fantasy Videophone Station completes the Fantasy Video Circuit. The two Fantasy Videophone Stations communicate information to each other through the Internet Cloud or POTS ordinary Telephone Network Cloud 1500.

Figure 15B illustrates one portion of this Fantasy Video Circuit. A user also acting as a viewer 0003 sits in front of a Fantasy Video Sender 0008 consisting of an Image Acquisition Device Mounted Above the Screen 0023c; a computer that is a family game play station acting as a "TV set-top device connected to a TV set or monitor" 0040l, which is used to support the Image Information Representation Subsystem 0021; and a Means for Making Information Available to the Distribution Channel that is a Telephone Wall Socket 1510. The Distribution Channel 0011 consists of the Internet Cloud or POTS ordinary Telephone Network Cloud 1500. The user also acting as a viewer 0003 is also sitting in front of a Fantasy Video Receiver 0009 consisting of a Telephone Wall Socket 1510 that embodies the Means for Accepting Information from the Distribution Channel 0017; a computer that is a family game play station acting as a "TV set-top device connected to a TV set or monitor" 0040l, which supports the Presentation Construction Subsystem 0030; and a connected television monitor 0040b. The Fantasy Video Sender 0008 together with the Fantasy Video Receiver 0009 compose a Fantasy Videophone Station 0007.

Figure 15C illustrates another popular embodiment of a Presentation System 0012 or an Imaging System 0010. The software is embodied on a PC computer 1550. The Presentation System 0012 consists of the PC computer 1550 running the software for the Presentation Construction Subsystem 0030; along with the Computer Monitor 0040a being the Presentation Device 0040. It is showing a Presentation of a User 0060. When used as an Imaging System 0010, the system also makes use of the camera Image Acquisition Device Mounted Above the Screen 0023c, along with the microphone Sound Image Acquisition Device 0023m, plus appropriate Means for Digitizing Images embodied as digitizing cards in the PC computer 1550. The software for the Image Information Representation Subsystem 0021 runs in the PC computer 1550. Adding an Internet connection would turn the system into a Fantasy Videophone Station 0007.

Figure 16A illustrates one embodiment of the Fantasy Video Receiver 0009, as combined with a multimedia presentation system. A presentation device 0040 presents the Presentation Combined With Multimedia Presentation System 1610. The multimedia presentation system can put up headings, display images and play sounds, and run other equipment such as MIDI synthesizers, curtains, or laser light shows 1611.

Figure 16B illustrates an embodiment of the Fantasy Video Receiver 0009, as combined with a hypermedia interface such as the World Wide Web or a local LAN-based hypertext system. It shows the Presentation Combined With HyperMedia Interface 1620 being presented on a presentation device 0040. So-called "hot spots" or buttons support jumping to different "pages". Different semantic or syntactic parts of the presentation, such as the presented hands or face, can also be hot spots.

Figures 16C illustrates an embodiment of the Fantasy Video Receiver 0009, as combined with a computer game. It shows the Presentation Combined With Computer Game 1630 being presented on a presentation device 0040. The game can be running on any one or several of the presentation devices discussed in Figs. 8A, 8B, 8C. User presentations 0050 can be separate or can be an integral part of the game.

ALTERNATE EMBODIMENTS

1A. Internet point-to-point Fantasy Videophone using computers

A typical embodiment of the Fantasy Videophone uses home computers that are attached to the Internet, to effect point-to-point bi-directional communication between a single user and a single viewer. Since all significant operation of the Fantasy Videophone is symmetric with respect to bi-directional usage being composed of two single-directional channels, of like or of differing levels of capability, here we only discuss communication in a single direction.

General Handshaking Steps. A Fantasy Videophone conversation typically consists of the following steps: (1) user imaging system calibration (2) user appearance and environmental appearance change (formatting information) specification by the user (3) establishing the distribution channel between the user's imaging system and the viewer's presentation system (4) user appearance and environmental appearance change (formatting information) specification by the viewer, if desired (5) initialization and negotiation of the actual formatting presentation construction information to be used (6) use of the Fantasy Videophone: (a) an image of the user and the user's environment is captured in a video frame and/or sound frame, etc. (b) the essential image information is represented in the imaging system (c) the representation is sent through the distribution channel to the viewer's presentation system (d) the presentation system creates a presentation using the essential information and the formatting information (e) the presentation system presents the presentation to the viewer; finally, (7) the connection is terminated.

Hardware/System Configuration. The **imaging system** 0010 in this embodiment physically consists of a powered color CCD camera; an output cable connecting the CCD camera to a video board; a video digitizing board inside the user's computer that accepts a video signal from the CCD color camera and captures it for computer use; a microphone; a sound board that digitizes the sound signal and captures it for computer use; the user's computer itself; and the software inside the computer that performs the functions of acquiring the image and representing the image information. In this embodiment, the powered camera constitutes an image acquisition device 0023; the cable plus the video digitizing board and driving software constitute a Means for Digitizing Images 0024; the microphone constitutes another image acquisition device 0023; the sound board constitutes another Means for Digitizing Images 0024; and the software constitutes the Image Information Representation Subsystem 0021. The computer, along with the digitizing boards it contains, represents the Means for Processing Images 0025; the camera and cable plus its digitizer is one Image Acquisition Subsystem 0020, as is the microphone plus its digitizer.

The **distribution channel** physically consists of a telephone modem inside the user's computer; a telephone wire that connects the telephone modem through a wall socket to the local telephone network cloud; the telephone network cloud itself; an Internet service provider (ISP) connected to the telephone network cloud and to the Internet cloud; a remote ISP similarly connected to the Internet cloud and to its local telephone network cloud; the corresponding remote telephone network cloud; a telephone wire leading from the remote telephone network cloud to the viewer's computer's modem; and the associated telephone modem connecting the incoming telephone wire to the viewer's computer; along with modem software on both the user's computer and also the viewer's computer to run the modems and effect communications. In alternative embodiments, the user's computer connects directly to the viewer's computer through a wire cable, or connects to it through the telephone cloud without using the Internet.

The **presentation system** 0012 is free to vary in kind in a manner completely separate from how the imaging system 0010 is implemented. Assuming we stay with the same technology for this example, the presentation system 0012 physically consists of the viewer's computer; a cable leading to a powered 2D color monitor; the monitor itself; and the presentation software necessary to run the system. The software embodies the presentation construction subsystem 0030; the monitor plus required support hardware embodies the Presentation Device 0040.

The modem software and software routines that use the modem software on the user's computer form the Means for Making Information Available To The Distribution Channel 0014; similarly, the modem software and software routines that use the modem software on the viewer's computer form the Means for Accepting Information From The Distribution Channel 0017. Thus, the described imaging system 0010 together with the modem software and software routines that use the modem software on the user's computer form a **Fantasy Video Sender** 0008, and the described presentation system 0012 together

with the modem software and software routines that use the modem software on the viewer's computer form a **Fantasy Video Receiver** 0009.

Altogether this forms a unidirectional Fantasy Video Circuit. If the user also has a powered computer monitor with a cable, and presentation software, then the user also has a presentation system 0012. If the user has in addition incoming modem software that can read signals from the modem plus software routines to use this incoming modem software, then the user also has a Fantasy Video Receiver 0009. This makes the user's hardware of the camera plus microphone, monitor, and computer stuffed with digitizing boards and a modem card, into a **Fantasy Videophone Station** 0007. If the viewer also has a similar set-up, then the two of them are free to establish a bi-directional Fantasy Video Circuit as is shown in the right half of Fig. 1D. The viewer could choose to only watch, however, not turning on her camera and microphone, which would make it only a unidirectional Fantasy Video Circuit, even though both of them have Fantasy Videophone Stations.

(1) User imaging system calibration. The Fantasy Videophone will want to separate the image portions of the user from those of the user's environment. There are a number of ways to do this; the most straightforward method is discussed here first. The imaging system first calibrates itself by asking the user to leave the scene. The imaging system takes a visual picture of the environment without the user, and optionally takes a sound sample of the ambient environment with the user remaining silent. Then the user is asked by the system to resume his or her place, by means of a sound signal and/or screen signal. It is expected that the camera and microphone will remain in the same place; if one of them is bumped, the user should ask the system to recalibrate.

Calibration depends upon the particular methods used in the software. An advanced system might want to take a picture of the user's head from different directions, in order to abstract a 3D model or a morph map. It might also take a voice sample in order to abstract a voice font. These calibrations could be stored and re-used by default, or re-taken each time the system is used.

A very advanced system would calibrate on the fly as the user starts using the system. The environmental image would be integrated from different pieces around the edge of the user as the user moves; the 3D model or morph model of the user would be integrated as the user moves around.

(2) User appearance and environmental appearance change (formatting information) specification by the user. The user must now select which enhancements are to be applied to the presentation constructed for the viewer. A simple embodiment uses only environment replacement, and transmits a literal image of the user and the sound. The user is presented with a menu of a set of standard background images; these can be photographs, computer-generated environments, images of paintings, etc. The user selects one for the current use. This selection code may be saved in a preference file on the user's computer for later default usage. A more complex embodiment could have a selection menu for which voice font the user wants to use today, or for which replacement costume, augmentation costume, or other

enhancements the user wants to use today. This would have to be a checked menu or a menu with X boxes in it to support combination enhancements in an advanced system.

The system would most likely compile this menu on the fly by querying any and all local or remote Libraries Of Formatting Information 0900 it might be able to find. Usually the user will have a few Libraries stored on a CD-ROM or on the local hard drive of the computer; there will also be others along the Internet.

(3) Establishing the distribution channel between the user's imaging system and the viewer's presentation system. The user asks the Videophone to form a communication connection between the user and the viewer. Alternatively, the viewer may have initiated the connection. The connection is formed in this case by use of Internet IP addresses and sockets. Given the IP address of the viewer, the user's computer opens a socket connection from the Fantasy Videophone imaging system in the user's computer to the Videophone presentation system in the viewer's computer.

In the case of a direct wire connection or a direct telephone-network connection, the transport layer for the distribution channel is a lot more simple. IP sockets might still be used, or a more direct computer-computer network link such as an Ethernet or Firewire pipe might be established.

At this point, the viewer is aware or is made aware that a Fantasy Video Circuit is forming. There are a number of different scenarios that could be supported:

(a) The viewer initiated the link to the user, who is already constantly broadcasting

(b) The viewer and user have already agreed, by telephone or by prearrangement, to initiate the link together right now or at a particular time

(c) The user initiates the link to an unaware viewer, and a "ring" signal is made to the viewer; the viewer is allowed to pick up or decline the call

(d) The viewer's Fantasy Video Receiver is always on, as is currently the case with fax machines. The user pushes a one-way or two-way conversation at the viewer.

(e) The user's initiation automatically turns on the viewer's Fantasy Video Receiver

(f) In the case of certain 911 or surveillance applications, the user's initiation can automatically turn on the viewer's Fantasy Video Sender

(g) The link is for a non-real-time e-mail application, and a computer picks up a message and records it for later spooling when the viewer chooses.

(4) User appearance and environmental appearance change (formatting information) specification by the viewer, if desired. The viewer now has the option of selecting what the viewer wants

the user and the environment to look like. This can be done using menus similar to those displayed for the user.

The imaging system acquires an image of the scene with the user present. The imaging system compares a video frame of the current scene with the image of the scene without the user. Lightness, color, and texture features are abstracted for each pixel in each image. Then each pixel in the current image is compared against its mate in the calibration environment-only image. If the features of the pixel all match those of its mate to within a prespecified feature-dependent threshold window, then the pixel is declared to be "unchanged", i.e., part of the environment. If any of the features differs significantly from those of its mate, then the pixel is declared to be "changed", i.e., belonging to the image of the user. This action is performed as fast as the user's computer can accomplish it, preferably at video frame rates.

Note that this method works for static environments, such as a home or office. It does not work well for moving environments, such as a bus terminal. In addition, it is useful to disable the automatic white balance on the CCD camera, as the background environment colors and lightness levels can change if the camera adjusts itself based on the presence of a user in the foreground.

(5) Initialization and negotiation of the actual formatting presentation construction information to be used. The user's imaging system now negotiates the environment presentation information with the viewer's presentation system. This consists of two parts: the selection, and making the formatting information available. As regards to the selection, there are a few conventions that are possible. The user can hold the final advantage ("trump") in some cases, the viewer can trump in some cases, or the last person who specified a particular change can trump in some cases. This is similar to the current-art problem of who trumps when a person wanting to hide their number calls a person or a police station having the caller-ID feature that displays the caller's number on their telephone. As for making the formatting information available, there are a number of alternative methods for performing this. Say, for sake of example, that we only want to replace the environment; selecting an algorithm, dataset, or list of multimedia materials needed for a different, more complex enhancement proceeds in a similar manner:

a) A number of standard environments are built in to both the imaging system and the presentation system. Environment-presentation information specification consists of a simple code number or a code word. The imaging system sends this selection code number or word to the presentation system during the initialization phase of the Fantasy Videophone conversation. Then the presentation system uses this selection code to index in to the selected standard environment. For instance, if there are five standard environments, entitled "BEACH", "CANYON", "SPA", "SPACESHIP", and "SALOON", then the user can choose "BEACH" and send this code to the presentation system. This process is similar to that of font specification for word-processing documents sent from a computer to another computer of a similar type.

b) The user's system has a custom environment that gets shipped to the viewer's presentation system during the initialization phase. This can come from a library of possible environments stored on the

user's computer. Or, it could be an environment that the imaging system creates for that session, perhaps with the interactive help of the user. The presentation information is sent from the user's computer to the viewer's presentation system by means of a distribution channel, perhaps a different channel than the one being used to support the conversation. For instance, an environment could be sent by CD-ROM or cable, while the conversation is about to be supported on Internet computer telephone (CT). The viewer's computer stores the presentation information received from the user's side and uses it during presentation construction. As an example, the user could send down a digitized picture of a particular beach to be used as an environment. This process is similar to sending down a custom font at the beginning of a document or sending it before the document, when the viewer's computer does not yet have the font.

c) A third-party entity can maintain a library of environments and costumes. The user can then query the library, make a selection, and send the selection code plus the third-party access information to the viewer's computer. The viewer's presentation system can then query the third party using the selection code and retrieve the necessary information for constructing the presentation. For instance, the user could specify "BEACH51 from Wild Animation". The presentation system would then query the Wild Animation server, download the photograph or model for BEACH51, and use that when generating the environment for this Fantasy Videophone call. This process is similar to referring to a font library on-line.

d) The viewer's system can have a library of environments and costumes. The user can then query this library after the distribution channel has been set up, make a selection, and send the selection code to the viewer's computer. For instance, if the viewer is famous for having an excellent Beach environment collection, then the user can instruct the imaging system menu to ask the viewer's computer for small thumbnail samples. The user can then select one of these samples. This selection code is then sent by the user negotiation component to the viewer negotiation component.

e) The viewer's system can have a custom library of environments and costumes that is indexed by description. The user's imaging system menu can have a similar catalog. If the user selects "A LONELY BEACH" for the environment, the viewer's presentation system can look up and substitute a similar environment with the same flavor from the library locally available to the viewer. The environment or costume will not be identical to that selected by the user, but it will be equivalent or similar. This is similar to the way HTML fonts are specified in 1998.

f) The viewer can override the user's selection with selections of his or her own, as previously discussed.

g) The appearance information could be generated from a random selection or from a selection based on the time of day, the day of the week, the next approaching holiday, or the queried or perceived emotional state of the user or viewer.

h) The appearance information could be assigned to the user's presentation by a third party. This will happen in tournaments, and when a director hands out costumes for a play or a program.

i) It is possible to have a combination of the above options. Thus, the presentation of the user's face could be generated from a local library, the body could be generated from information taken from a third-party server, and the environment could be sent directly from the user's computer on initialization.

(6) Use of the Fantasy Videophone. At this point, the Fantasy Video Circuit has been established, and the viewer is ready to being receiving presentations from the user.

As was stated, this breaks down into: (a) an image of the user and the user's environment is captured in a video frame and/or sound frame, etc. (b) the essential image information is represented in the imaging system (c) the representation is sent through the distribution channel to the viewer's presentation system (d) the presentation system creates a presentation using the essential information and the formatting information (e) the presentation system presents the presentation to the viewer

(a) **an image of the user and the user's environment is captured in a video frame and/or sound frame, etc.** This is done using standard technology, by using the image acquisition device(s) and their corresponding means for digitization. For instance, in the case of a video camera with a digitizer, the system grabs the next available video frame.

(b) **the essential image information is represented in the imaging system.** This is a complex step. Two proposed technologies are discussed in detail after (7). In most systems, the sound will be represented literally. In simple systems, the image of the user extracted from the environment 0270a might be used for the essential information and represented literally as well, or the system might trim out and represent literal images of only the user's face, or eyes, eyebrows, and mouth. More advanced schemes are discussed below.

(c) **the representation is sent through the distribution channel to the viewer's presentation system.** This step is quite straightforward. In the case of a TCP/IP pipe, for instance, the imaging system divides the essential information into buffers that are then sent to the lower levels of the computer's operating system. The lower levels decompose the buffers into packets as necessary. These packets are shipped out over the pipe, captured on the other side, reassembled by lower levels of the viewer's computer's operating system, and presented in buffers to the presentation system. This is a standard problem describing application communication over a transport layer, which must be solved for most modern forms of telecommunication technologies before they can be used as distribution channels.

(d) **the presentation system creates a presentation using the essential information and the formatting information.** This step is coupled to step 6(b), and is also discussed below. The output is a buffer of digital information suitable for perception by a viewer, such as a sound data buffer or a buffer of video frame data.

(e) **the presentation system presents the presentation to the viewer** This step is done using standard technology, by using the presentation device in its intended manner. The buffer of digital information comprising the internal presentation is sent to the presentation device, at which point it is made available for sensory consumption by the viewer. In the case of a video system, typically this is done by writing to or constructing the buffer in video memory, which maps to a display on the connected computer monitor. In the case of a sound system, typically this is done by sending the buffer to a sound routine that inserts it into a hardware buffer, which gets written out to the one or more speakers by means of a DAC (Digital-to-Analog Converter). In this manner, the viewer gets to perceive the presentation.

(7) The connection is terminated.

Termination in the case of a non-billing point-to-point situation is straightforward; one party or the other simply exits the software or turns off their computer. In the case of a toll call being billed through a network, it is necessary to send a termination signal back to the network billing authority at this point, so that billing time does not continue to be metered. In the case of a multicast or broadcast situation, the broadcasting unit might have to receive a signal that one of the viewers has decided to terminate, and might have to make a decision as to whether to continue sending signal in that direction, or whether to collapse that portion of the distribution channel if there aren't any other viewers in that direction.

1B. Fantasy Videocircuit Using a Computer Graphics/Robotics-Based Method for Representing and Presenting Essential Information.

We now turn to the problem of how to abstract and represent the essential information used for communication, and its corresponding problem of how to reconstruct a possibly enhanced presentation, using the essential information and formatting information. There are two main embodiments proposed for these steps for visual presentations: one based on computer graphics/robotics methods, and one based on methods using perspective morphing of 2D images.

In the preferred embodiment, the image information representation subsystem and the presentation construction subsystem use the following computer-graphics/robotics algorithms. This method is illustrated in Fig. 3A. The image information representation subsystem uses a software visual face tracking

system and body tracking system to keep track of the location (that is, the six-degree-of-freedom position and orientation) of key parts of the user's face, head, and body, relative to the user's environment. (This could also be done using tracking hardware, which works more rapidly, but is inconvenient and more expensive.) For instance, the system tracks the outside and inside corners of the eyes, and uses the relative distances and angles between these points to determine the location of the head. The size of the head image and the width of the shoulders helps to determine distance from the camera in a 2D system; the top part of the head image determines the height of the user, etc. Then the subsystem abstracts a global location for the user, along with relative joint angles and muscle actuation values for each essential joint and muscle. For instance, the amounts of elbow bend, eyebrow lift, jaw opening, and lip pucker should be abstracted and recorded in a representation data structure. Some parameters, such as eye squint factor, will describe a local configuration and not just a local activator position. The subsystem records an actuation value for each significant joint or muscle in a standard human body model, which is mapped onto the user. Working with the body, including arms, legs, the torso, neck, and fingers, is relatively straightforward. One set of actuation variables for the more difficult face can be found in "Facial Action Coding System" by Ekman and Friesen, Consulting Psychologists Press, Inc, 1978. The subsystem records this information as fast as is possible and as fast as is needed. Typically any rate over 30 frames per second is wasted effort.

The subsystem also uses visual information to abstract the location, extent, and frequency spectrum of sources of illumination (lights) in the scene. This can be done by assuming that the user is a constant color, that the lights are also of constant color and location, and that any changes in the scene's colors are caused by the interaction between the user and the light sources as the user moves. It is possible to track patches of the user's appearance as the user moves. An initial default scene-illumination model is proposed, and then used to explain variations in a patch's color. The color of the patch is integrated over time and ranges are observed. Unexplained variations in patch color cause revisions in the lighting model and patch color model. In this way, both an illumination model for the environment and a so-called "texture" model (actually a color plus texture model, taken from the patches) for the user and for the environment are abstracted.

The subsystem also uses aural information to abstract the characteristic speech sounds of the user, such as phonemes, current emotions, loudness, and pitch. These should also be recorded at around 30 frames per second. The subsystem also abstracts a "voice font" that describes the standard voice of the user through its range of colorations.

At this point, as previously discussed, negotiation occurs between the user's side of the system and the viewer's side of the system, as to which environment and appearance change information will actually be used by the viewer's presentation system. This will typically have a 3-D environmental model, along with a virtual costume for the user's body, a voice costume for the user, and instructions for replacing the user's body and voice. The virtual costume has a spatial model, a texture model (including colors, bump

maps, and reflectance maps), and an actuator effect model ("wiring"). The environmental model contains similar information, except that it typically is assumed to be stationary and thus does not require an actuator effect model in this embodiment. The environmental model will also typically have camera lighting models associated with it. Negotiation occurs, as previously discussed, by the user's imaging system suggesting a source location and content code for the change information, and by the viewer's presentation system either accepting the suggestion, denying that such information is accessible and requesting another round of negotiation, or overriding the suggestion with its own preferences. As discussed, the information source can be located on the user's side, on the viewer's side, or at a third party site, and can be distributed; it doesn't have to be all in the same place. After the negotiations are concluded, the presentation formatting information is downloaded (if necessary) and made available.

For parts of the user and environment that are not going to be replaced or deleted, the image information representation subsystem sends down a relatively large initial download through the distribution channel to the presentation system, after the image information representation subsystem has been calibrated but before the actual conversation or broadcast begins. This information has geometric models of the user and relevant environmental objects; texture models of the user and the environment; a lighting model of the scene; and a voice font of the user. This information is assumed to be static. It can be re-downloaded when a change is detected, or on a periodic basis to ensure synchronization, depending upon the length of the conversation or broadcast.

The joint and actuator information, along with the characteristic speech sounds, is dynamic information that should be distributed in real-time for online embodiments.

The presentation construction subsystem 0030 takes the static and dynamic information, along with the change instructions, and creates a presentation. The dynamic pose information is used, through the actuator wiring information, to reset essential information in the formatting model of the user, such as the elbow joint angles, mouth pucker, and user location, etc. into a new pose. Then the formatting model of the environment and the positioned model of the user are rendered into a presentation. See "A Collision-Avoidance System for Robot Controllers", Myers, Master's thesis, Carnegie-Mellon University 1981, for further information on this now-standard task.

Flexible objects such as hair and clothing, and liquids, are handled by a dynamic simulator. The imaging system abstracts enough essential information as to be able to characterize the objects. This information is sent to the presentation system. The presentation system maintains a dynamic simulator for the set of objects, and then modifies the animation as instructions are received.

Besides standard polygonal techniques, the environment or user model presentation information can use NURBS, metaballs, QuickTime VR models, light flow fields or lumigraphs, etc.

Sound is rendered in a similar manner. The chosen voice font is used, along with the characteristic speech sound information, to re-create or to create a new appearance for the speech of the user.

Other sensory perceptions such as range, touch, force, taste, or smell can be handled in a similar manner with similar sensory imaging systems, change mappings, and presentation systems.

1C. Fantasy Videocircuit Using a Perspective Warp Method for Representing and Presenting Essential Information. This method is illustrated in Fig. 3B. The imaging system acquires raw images from the user. A list of visual features is maintained and tracked, based on texture, color, outlines, corners, etc. Each feature has a two-dimensional point location in each image in the video stream. The features are back-projected into the user's 3-space, using spatial and temporal coherence information to resolve missing information. Then the virtual camera is moved. The new 2D location of each feature point in the current video-frame image or in the image of a costume is computed for the new location of the virtual camera. Then a perspective morph is performed on the input image or costume image. The perspective morph takes as input the image of the user or the image of a desired costume; a list of 2D point locations in the image, as seen from the actual camera or the costume-image camera, corresponding to the locations of tracked features; and a list of desired 2D point locations in the presentation image to be constructed, as seen from the virtual camera position. The perspective morph stretches and overlaps the image according to the point locations, and gives as output the morphed video frame image corresponding to the view of the user or user's costume as seen from the virtual camera position. Similar methods can be used to composite in augmentations. [See "View Morphing" by Seitz and Dyer for more information].

1D. Fantasy "Video" using Sound Images

Although the term "Fantasy Video" has been used to describe the system for convenience, because it is envisioned that most of the popular enhancements will be performed in the visual modality, there is nothing restricting the process to visual images. Another important modality is working with sound images.

In this case, the Image Acquisition Device is a microphone or set of microphones, and the Means for Digitizing Images is a sound card that can take microphone inputs.

Separating the speech of a user from background environmental sounds is a difficult task and one that is still being researched. If the environmental sounds are consistent in their frequencies and outside the main power spectrum of the user's voice, it is possible to ask the user to be silent for a minute, record the environmental sound image only, form a power spectrum of its frequencies, ask the user to speak, form a

power spectrum of the environment plus the user, subtract the two to get a typical power spectrum of the user, and then use the typical power spectrum of the user and the typical power spectrum of the environment as inputs to a maximum-entropy recognizer that attempts to label image frequency components as belonging to user or environment.

It is probably easier in this case, however, to use an engineering solution. Noise-canceling headset microphones work wonders.

As a side note, one of the capabilities of a visual Fantasy Video Circuit is to perform a "clean up" or "deletion" enhancement and remove the image of a headset of a call-center operator from the presentation of a viewer. Alternatively, call-center operators using focused-microphone speakerphones can have headset microphone accessories added to their appearance using an "augmentation" enhancement.

Returning to the sound image system, once the speech of a user has been isolated from environmental sounds, its essential information is analyzed and represented by the Image Information Representation Subsystem.

One embodiment of this uses phoneme identity, duration, loudness represented as total power, and pitch represented as fundamental frequency for its essential-information paradigm. A phoneme-recognition system, which is a common component as the bottom layer of a speech-recognition system, segments the speech into phonemes of varying lengths. Then an analysis is performed on each segment, to measure its total power (energy), its duration, and its fundamental frequency when appropriate. The phoneme identity is recognized and codified. Then the identity, duration, power (loudness), and fundamental frequency (pitch) are declared to be the essential information, packaged into a structure, packed into a buffer along with other such structures, and made available for delivery to a distribution channel. This constitutes the Imaging System side of things.

The Presentation System runs the system in reverse. The Presentation System has formatting information in the form of a voice font, either of the user or of someone or something else. Different voice fonts can be used, and the user can switch between chosen voice fonts at will for different presentation effects.

One embodiment of a voice font has a range of phoneme sounds that have been recorded at differing frequencies. The Presentation Construction Subsystem then selects which recording to use for each temporal segment, based on phoneme identity and frequency. The phoneme sound is scaled for loudness. Then the phoneme is tuned for duration. If the recorded duration is longer than the specified

duration, it can be truncated; if the recorded duration is shorter than the specified duration, the last half of the phoneme recording can be replicated for as long as required. More elaborate schemes involve using starts, middles, and ends of phonemes, where the middle's duration is tuned as required.

Finally the tuned phoneme sounds for each segment are concatenated together. An abrupt triangular fade-in and fade-out should be used for the start and end of each segment sound, and the segments should be concatenated in a slightly overlapping fashion, to avoid too-abrupt changes between adjacent segments.

In this manner, a Fantasy Video Circuit can communicate vocal images that are subject to enhancement as well.

1E. Fantasy Video TV/Movies

Although many people will want to use a two-way Fantasy Videophone Station, many other people will not be interested in communicating with someone else in a two-way fashion but will only want to view a presentation made by someone else in a one-way fashion. This is called watching Fantasy Video TV/Movies, because it is similar to watching TV in 1998, or to watching movies in a movie theater or short QuickTime™ movies on the Web. The viewer operates his or her machine as a Fantasy Video Receiver. In this case, a camera or microphone and corresponding digitizer is not necessary; the viewer only needs a machine with a presentation device (see Figs 8A, 8B, 8C) attached to a computer processor running the Presentation Construction Subsystem software and connected to a distribution channel such as a telephone wire, an Internet feed, or a television-band radio-waves receiver. The viewer turns on the Fantasy Video Receiver and selects an interesting "TV channel" to watch. On the Internet, this will probably be referenced by an IP address of a Fantasy Video TV Broadcasting company; or, the user might have a special telephone number to dial to join a broadcast; or, the program could be broadcast on cable; or, certain parts of the radio or television spectrum might be devoted to broadcasting Fantasy Video programs. The viewer's Fantasy Video Receiver ties in to a stream of information that has already been set up. The stream specifies the dynamic action (movements, speech) that is going on in the program; a second stream, which might be interwoven with the first, sporadically specifies the static parameters for the costumes, environments, and enhancements required for the presentations. The viewer's Fantasy Video Receiver accepts these streams, creates an ongoing presentation, and sends it to the Presentation Device(s) for the enjoyment of the viewer. The viewer can switch "TV channels" at any time, or instruct the presentation system to use different costumes or environments of the viewer's choice by means of a menu. So, for instance, the viewer can place an avatar of him or herself in the story to replace the appearance of the main

character, can insert a voice font of his or her own voice in to likewise replace the vocal characteristics of the main character, and can insert an avatar of the viewer's favorite sex symbol in to play the love interest character. More creative options are possible.

Since a Fantasy Video Receiver unit for Fantasy Video TV does not have to have the slightly more expensive camera and digitizer included, it can be produced and sold for slightly less than a comparable Fantasy Videophone Station. The form-factor for such a device has little to do with its design; any of the Presentation Devices shown in Fig 8A, 8B, and 8C can be used in a successful manner for the front-end output device for a Fantasy Video TV Receiver. It is merely necessary to have the proper driving electronics and video buffers or sound buffers in the computer or processor being used to support the Presentation Construction Subsystem. For instance, the HDTV device 0040c is going to require significantly different driving electronics and video buffer space than the hand-held game device 0040y. In addition, the viewer might have to adjust the picture by adjusting the lens angle and position of the virtual camera being used.

One of the fascinating advantages of the Fantasy Video system is that output Presentation Devices of all different resolutions and capabilities can be used for viewing the same essential-information signal. Since the presentation is constructed by the Presentation Construction Subsystem, which must be aware of which Presentation Device(s) it is handling, the presentation can be made especially for that device. For instance, a vertical device such as the tall building sign 0040h will have an aspect ratio that is quite different from a horizontal device such as the HDTV 0040c. However, the respective Presentation Construction Subsystems driving these devices can fill in more of the central actor's body in the case of the vertical aspect ratio, or more environment or more actors in the case of the horizontal aspect ratio, and thus display differing bits of more of the scene but still display the same scene for these two wildly-differing presentation devices. The same Presentation Construction Subsystem can even handle both devices at the same time, if the Subsystem has enough memory and is fast enough to be able to jump back and forth between them.

When literal textures are used, the Presentation Construction Subsystem can interpolate the textures for output devices of higher resolution than the literal texture, or low-pass filter or decimate the textures for output devices of lower resolution.

The invention also supports devices of differing temporal resolutions. Current-day TVs require 30 frames per second to be displayed, whereas current-day movies require 24 frames per second. Some monitors can handle 60 frames per second or more, whereas some low-end computers can only generate 20 frames per second. The Presentation Construction Subsystem should keep a clock running so that it times how fast it takes to construct each frame. It should also know the top frame rates of the Presentation Device(s) that it is using. The Subsystem should try to match its construction frame rate with the fastest

supported rate of its output device. If the Subsystem is slow, it can make trade-off decisions as to how much level of detail to put into constructing presentations, in order to get a faster frame rate.

The invention can even interpolate in time if the presentation device is significantly faster than the rate at which the originating imaging system sampled the scene.

In this manner, a viewer can watch Fantasy Video TV on any Fantasy Video Receiver system.

1F. Fantasy Video TV/Movies Prerecorded Web Browser

Because the essential-information stream is so light-weight, it can be streamed over the Internet with little difficulty. Users can record a stream of essential information into a file, by using an outgoing message recording system and a storage system that is a file on a hard disk. A user can then review and edit the Fantasy Video information in the file. The user can then post the file onto his or her web page. The file contains all of the essential information for displaying the Fantasy Video TV/Movie, in a manner similar to Apple's QuickTime™ format. The file can include formatting information for instantiating environments and enhancements; it can contain codes or pointers to well-known formatting information; or it can leave the formatting information up to the default used by the viewer.

A Fantasy Video presentation can be combined with other multimedia interaction to form a Fantasy Video hypermedia system. Titles and text can appear in 2D or 3D in and around the Fantasy Video presentation. 2D pictures can be displayed on top of, in, or behind the scene; music and other sounds can be mixed in. All kinds of outputs can be combined. Inputs can be combined, too, by making different parts of the scene selectable or click-able, and putting URLs or other "hot links" in the action to be executed when a part of the scene is selected.

1G. Example Enhancements

Enhancements may consist of such transformations as adding or subtracting make-up, modifying skin color, modifying eye, hair, or lip color, modifying pupil diameter, modifying facial features communicating such signals as tiredness, interest, happiness, authority, etc.; modifying apparent age; modifying apparent gender; making the user fatter or thinner, making the user taller or shorter, making the user more or less muscular, making the user have more or less body fat, breast size modification, shoulder size modification, hair length modification, making the user have a different haircut, adding hair or subtracting hair from the head, face, or body of the user (also including mustaches and beards), modifying teeth straightness, pointiness, or length, modifying facial features and deformities, modifying eyeglasses or their lack, adding extra features, including for example a third eye, horns, antennae, bug eyes, a tail, horses' legs, extra legs, an extra mouth, etc. adding or subtracting tattoos; adding a halo, an aura, a shadow, or other modifications to the lighting in the scene; rendering portions of the user invisible, transparent, or translucent; modifying the apparent visual construction of the user, by making the user seem as if the user

is made out of metal, plastic, leather, or other material; adding feathers or wings; adding scales, hide, or different kinds of skin such as toad or cow skin; adding chitin; adding cyborg implants such as cameras, tubes, and hardware; adding accessories such as a hat, jewelry, piercings, accessories, extra clothing, high heels, etc; or making the user seem to levitate. These are examples of enhancements that can change the appearance of the user's presentation. Other enhancements are possible.

1H. Multiple Users and Multiple Viewers.

There are a number of ways that multiple users can use the same imaging system. The most straightforward way is to have each user be assigned to one or two imaging devices, and then send each separate scene down in parallel. The viewer can then combine these with a split screen or with mini-screen windows; the viewer can have a "picture-in-a-picture" large window with tiny overlay windows at the bottom acting as icons to click upon; the viewer can have multiple presentation devices; or the viewer can have a screen that flips back and forth between the user presentations.

The second method is to assign each user one or a few image acquisition devices, and have the Imaging System or the Presentation System combine these together into one environment, so it looks as if the users are all together in one scene.

The third method is to have multiple users per scene in front of the image acquisition device(s). In this case, the Image Information Representation Subsystem must track the individual users and separate them into separate environmental scenes or leave them together, perhaps repositioning them, in the same scene.

It is of course possible to have multiple image acquisition devices and multiple users in front of some or all of these devices.

In any case, each user can specify particular enhancements for his or her own presentation in a separate manner, or use blanket changes all together. A viewer can also request specific enhancements for each user's presentation and for each environment.

The case of multiple viewers is more straightforward. Several viewers can obviously use the same presentation device; or, multiple presentation devices can be hung off of one Presentation Construction Subsystem, each under central control or each under the control of a separate viewer; or, a combination of these.

2. **Regular point-to-point telephone network using a computer processor built into Fantasy Videophone Station standalone units.**

A standalone unit, that is, one that is not connected to a separate TV or computer, will come in a number of different form factors but they all will have similar methods of construction and operation. The most popular form factors for a standalone Fantasy Videophone Station are expected to be the so-called "wearable computer" 0040u, consisting of a small processing unit kept in a shirt-pocket or on a belt, along with a small screen that drops down in front of one or both eyes off of a headband; the cellular videophone 0040n, being a pocket phone that is carried around; the wrist-watch videophone 0040p; the desktop videophone 0040k1 or wall pay-videophone 0040k2; and the wall-mounted flat-screen 0040d videophone with an integral camera or two installed above it. In each case, a camera and digitizing electronics must be provided for the Imaging System hardware, along with a computer processing chip to take care of running the Image Information Representation Subsystem and Presentation Construction Subsystem software. The formatting information algorithms and associated multimedia elements can be standardized and burned into ROM, or downloaded from a computer-based Fantasy Video Sender if the standalone unit is receiving a call, or kept on a matchbox hard-drive and updated periodically, or downloaded from the telephone company (see Fig. 9C) as an extra service. The standalone unit itself requires typical telephone hardware and firmware in order to be able to establish, maintain, and break telephone calls. In half of the cases, this will also include a jack for a wire going out the back to go to the local telephone company or a private network; in the other half, this will also include an antenna and associated hardware and firmware to handle supporting cellular telephone circuits. In addition, the unit requires modem chips of some kind to be able to handle digital communication between the local videophone unit and whatever is on the other end of the Fantasy Video Circuit. These will be communicating using digital signals; either a high-quality digital telephone circuit must be established, requiring digital send/receive hardware, or a low-quality analog telephone circuit must be established, in which case the unit must turn its digital signals into analog tones using the modem chips. In any case, the Image Information Representation Subsystem of the local unit is able to communicate digitally with the Presentation Construction Subsystem of the remote unit, and vice versa. It is also necessary to build in a video screen that is hooked up to a video driver and video RAM accessible to the computer processor chip's memory. One of the main challenges will be providing power to the unit to drive the display and the processor; this will probably be accomplished with rechargeable lithium batteries in the near-term, and micro-turbines or sugar-burning cells in the long term, for the portable units; installed units can use wall power. Thus, the hardware for the unit will probably be able to use off-the-shelf components, but the configuration and circuit boards will have to be custom-designed and manufactured.

To recap, in this embodiment, the imaging system and the presentation system consist of videophones with smart processors built in. It is not necessary to use actual computers, merely dedicated custom-built computer hardware that handles the functions of the imaging system and the presentation system in this embodiment. The imaging system is physically a videophone with a small electronic camera built in, along with special-purpose hardware and software for managing telephone dialing, telephone

connections, image acquisition, and image representation, manufactured in a manner physically similar to existing videophone units. The presentation system is physically a videophone with a color imaging screen attached, along with similar special-purpose hardware and software for managing telephone connections, presentation construction, and displaying the presentation. In practice, video conversations will typically be two-way, and the hardware will be replicated on both sides. The only difference between current videophone technology and the embodiment is the special-purpose hardware and software necessary to support the functionality of information representation, presentation negotiation (if any), and presentation construction. Current technology handles both image acquisition and presentation display. Preferably, the distribution channel is the telephone network. Alternatively, this can be local, long-distance, or a local exchange located inside a single company, among others.

A user punches a button or picks up a handset to get the attention of the Fantasy Videophone. The Videophone presents a menu on the screen of a number of formatting possibilities, and the user selects both an appropriate environment and user-appearance changes, then hits an "O.K." button. The videophone should also give the user a "use defaults" button that simply uses the enhancements that the user chose the last time, without having to go through a menu. Upon hitting either of these buttons, the menu collapses and the videophone gives a dial-tone. The user dials a number on the standalone Fantasy Videophone. The standalone unit uses the telephone network to connect with another standalone Fantasy Videophone (or even with a computer telephone exchange providing a bridge to a personal computer Videophone as described in the first embodiment). The two units initialize, and negotiate the presentation information. The user's imaging system captures the user's image, represents it, and sends it across the telephone network to the viewer's presentation system, where it is presented to the viewer. The system should be able to handle 30 frames per second. Sound information is typically sent across in the same distribution channel; the sound can be represented using essential information based on phonemes, or it can have the user's voice extracted from the environmental sound, or it can simply be left alone and transmitted on another time-multiplexed band in the circuit.

When the person is viewing a call coming in, that person can have the option of modifying the appearance of the presentation of the user. Often the viewer will not choose to exercise this prerogative, but if the viewer wants to make the day more exciting, or objects to an obscene appearance of a user, the viewer can pull down a menu and modify the enhancement parameters of the presentation.

A wrist-watch videophone or a cell-phone videophone will have problems with the angle of use, as is illustrated in the left half of Fig 14. The user will most probably want to restage the presentation by moving the virtual camera to a spot four feet away and at eye level of the user, as illustrated at the bottom of Fig 14. This gives the most friendly appearance for American viewers. As long as the user's appearance is generated using a 3D model, the mathematics for doing this are almost trivial; the computer

graphics routines simply require a location for the camera point, which is moved to the desired position and orientation.

If the user's appearance is generated by abstracting the user out of the environment, shipping the literal textures of the user image, and then applying a transformation for display, restaging becomes more challenging. In this case, simple systems will apply an affine transformation or a perspective transformation to the user's video image; more advanced systems will attempt to apply a perspective warp by tracking key points of the user's face, such as the tip of the nose, the sides of the nose, and the hairline, and warping these in the single (moving) 2D user image to their proper locations based on a 3D model of these points' positions as determined from the desired restaged virtual camera. It is necessary to use key points when doing a perspective warp in order to ensure the warp comes out properly. It is necessary to use a perspective warp instead of a regular warp in order to ensure that the resulting 2D image looks as if it had actually come from a real 3D object, instead of being merely a squished photo. A "warp" of a single 2D image is the first step in constructing a morph between two 2D photographic images; in a morph, the first image is warped, and the second image is warped, and then a weighted average is taken of the two warped images to yield a morphed photographic image. Reference [3] Seitz and Dyer '96 discusses the new mathematics behind perspective warping and perspective morphing in great detail.

3A. Multi-user cyberspace maintained by third party network systems.

In this embodiment, a third-party company maintains a set of **environments** called a "cyberspace". These will typically consist of geometric models, color/texture information, and lighting information, although they could be QuickTime VR(tm) spherical or cylindrical photographs, or 4D or 5D light fields [8] Levoy & Hanrahan, etc. The environments could be business offices, grand hotels, dinosaur jungles, space asteroids, barbarian wastes, etc. The third-party company also will usually have a selection of **virtual costumes** and other formatting information for users to choose from. These could be cyborgs, razor-girls, barbarians, business suits, etc., for the costumes, and various augmentations (such as hats, swords, gold coins) and other enhancements for the other formatting information. The company maintains a computer server that provides telecommunication services. A user dials in and is instantiated in a virtual place in cyberspace, using a costume called the user's "avatar". One or more users can appear in the same environment, and can be seen by each other. The Fantasy Videophone is an improvement over current cyberspace worlds because the avatar reflects facial expressions and movement commands sent by the user in real-time.

Another interesting point is the virtual camera position. In this embodiment, viewers may choose a first-person viewpoint, an over-the-shoulder viewpoint, or a remote-camera viewpoint. When the viewer is also a user, the viewer/user may see the side of his/her body or the back of his/her head.

In this embodiment, each user will also typically be a viewer. Multiple users and multiple viewers will use the same distribution channel and connect to the same third-party server. There may be more viewers than users—some viewers will choose to "lurk", i.e. to be spectators and control a virtual camera into the cyberspace while watching on their Fantasy Video Receivers, but to not operate a Fantasy Video Sender.

This model is called a "star topology", because there is only one or perhaps a few central servers that provide information to multiple viewers, which conceptually surround them like rays of a star. See 0730 in Fig. 7.

3B. Cyberspace in a mesh or ring

Another embodiment circumvents the need for a central server. Fantasy Videophone users create a multi-user cyberspace by using a distribution channel that has more than one connection per Videophone. The users can connect in a mesh topology (0720), where each user's Videophone opens a connection to all of the other Videophones that the user wishes to communicate with. In this case, information is sent directly. Or, it is possible for each Videophone to accept a connection from only one caller, and to send information to only one viewer, but the caller and the viewer are typically not the same. The Videophone forwards all received information along, and adds information from its user; however, it subtracts any incoming information from the caller about its user and does not forward that, to avoid infinite loops. This is called a ring topology, since the information gets forwarded around in a ring (0710). The topology of the distribution channel does not matter to the central core of the Fantasy Videophone concept, since the user can present a changed appearance or a changed environment to any number of viewers.

3C. Small-party private cyberspaces maintained by third party network systems

It is not necessary for a cyberspace network company to support one huge cyberspace in which everyone who logs in participates. Commercial cyberspace companies can also support small-party private cyberspaces, where a small party is defined as a private group consisting of one, two, or a small handful of people. In this way a person and all of his or her friends can meet together and can enjoy exploring a cyberspace in a discreet manner.

3D. Cyberspace on a CD-ROM

It is not necessary to maintain a cyberspace as a separate central server. A cyberspace company can put environments, costumes, and enhancement algorithms on a CD-ROM or up on its web-site, and then sell them to customers having compatible Fantasy Video Receivers. The customers choose which components to use.

4A. Fantasy Video Email Movies

It is not necessary for the transmission over the distribution channel to occur simultaneously with the image acquisition and representation. In the Fantasy Video Email embodiment, additional components are added to the Fantasy Videophone design. A recording unit takes the information coming out of the image information representation subsystem and records it to a buffer in memory or a file on a recording device such as a computer disk. After the user is finished recording a message, an option on the control system asks the user whether the user wants to review, re-record, or send the message. A local presentation system on the user's machine allows the user to view the message, by accepting input from the buffer or from the recorded file instead of from a distribution channel. The user thus gets a chance to review the message, and see if it actually is what the user had in mind. A second option allows the user to delete the recorded message and re-record. On a bare-bones system, these options might not exist, and the video email might be sent directly. In any case, when the user is satisfied with the message, the control system invokes a sending module to copy the Fantasy Video Email message over the distribution channel to the viewer's machine. The message can be sent using regular email sending/receiving technology.

After the message has been received, the viewer invokes a special presentation system to watch the message. The presentation system accepts the message from the email as the distribution channel. The presentation system can be a stand-alone program, or it can be software plug-in for other email readers such as Netscape(tm)'s browser.

Similar to the Fantasy Videophone, the Fantasy Video Email presentation system requires negotiation as to how the message will be presented. This can come from the user and be bundled with the email message, can come from a reference to a third-party library, can be overridden by the viewer, etc., as previously described.

Of course the message can be sent with associated multimedia, such as sound, music, and other presentations. These can be displayed in a combination presentation system that uses appropriate associated technology to display both the Fantasy Video presentation and the associated multimedia simultaneously. For instance, a sound channel would require a sound presenter; a music channel would require a music presenter, etc.

4B. Recording and Editing a Fantasy Video TV/Movie with playback, splicing, laying down tracks

In order to be able to compose Movies for use on the Web or for broadcasting over a Fantasy Video Broadcasting Station, it is important to be able to edit the Fantasy Video TV/Movies. This is done using a local tool analogous to Adobe's Premiere™ editor that edits QuickTime™ movies. The editing tool should ideally have a recorder built into it, consisting of an Imaging System hooked up to a Recording System or Means For Recording that uses a Storage System. The Means For Recording is a simple piece of software that streams an outgoing stream of essential information into a file on a hard disk instead of out to the distribution channel; the Storage System is the hard disk.

The editing tool should have playback capability. This is accomplished by having a Presentation System to display the results of the file to the editing viewer, along with a Means For Message Playback that is a simple piece of software that opens an essential-information file and streams it into the Presentation System instead of having the information stream come from an outside distribution channel. Streaming is accomplished by reading information into a ring buffer from one side while simultaneously copying it out to the other side; semaphores are used to make sure overwriting does not occur. In this manner an editing viewer can view the presentation contents of a file.

The editing tool should have splicing capability. This is accomplished by each frame of information having a relative time-stamp in the stream of information. Then the editor can read in two Fantasy Video Movies; it can reach in and cut frames out from one time up through another time in one movie, and paste them in after the end of another movie. It can paste the information in one movie over the top of the information in another movie, by deleting the contents of the pasted-to movie during that duration. And, it can composite or "lay tracks" down in a movie by pasting the essential-information contents of one movie into the contents of another movie, but not delete the contents of the pasted-to movie during that duration. This is needed in cases in which one actor is recorded, and then a second actor is recorded, and the two of them are supposed to be doing a scene together. A good editor tool will keep tracks separate and will be able to play them simultaneously, without doing a composition, until instructed to. This is done by the editor maintaining "current-time" pointers into each movie track. The current time moves forward in each track as the multi-track composition being edited is played back, and the Presentation Construction Subsystem accepts information from each track and puts it together in real time as the presentation is being constructed. This allows the relative timings of each track to be adjusted on the fly. Then, when everything is finally made right, the tracks are "laid down" by performing a composition. Since composition takes each frame from each input movie and interleaves them to form a single output movie, it is a relatively time-consuming operation that could take longer than real-time in slower computers. For this reason, and to maintain flexibility, composition is typically not performed until the editing has been finished.

4C. **Fantasy Videophone answering machine.**

A user may record a Fantasy Video answering message to be played back when someone else calls and the user is not in. The Fantasy Videophone answering machine can then record a Fantasy Videophone message from the caller. This system is similar to the email system, except that the recording is done on the side of the called person, instead of the side of the caller.

In this embodiment, an answering-machine control component is notified when there is an incoming call. If the called person does not pick up, the answering-machine control component invokes an outgoing playback message and then, when the outgoing message is finished, a recording unit. The

outgoing playback message is stored in an outgoing-message information storage system, such as a hard disk on a computer, a Fantasy Video information tape, etc. The outgoing message is created by a Fantasy Video message recording system. The message recording system prompts the user to begin recording the message, and then intercepts the information stream that would normally go from the imaging system to the live distribution channel. In this case, however, the distribution channel consists of the message recording system, the outgoing-message storage system, and the answering-machine control component, along with the telephone network or Internet network used to send the message out. When it's time to play back the outgoing message, the control component accesses the outgoing-message information stored in the outgoing-message storage system and sends it down the distribution channel to the caller.

After the outgoing-message information has finished being sent, the control component then sends a request to create a recorded message to the caller. The caller uses the caller's imaging system to acquire and transmit a stream of Fantasy Videophone information through the distribution channel to the answering-machine controller. The answering-machine controller sends this information to an incoming-message recorder, which may be the same as the outgoing-message recorder. The incoming-message recorder records the caller's Fantasy Video message into an incoming-message storage system.

When the called person comes back, the called person queries the answering-machine control component as to whether there were any new calls or not. The person selects one or a couple calls from the set of new or old incoming messages. Then the answering-machine controller sends the information corresponding to these messages from the incoming-message storage system to the called-person's presentation system, where the Fantasy Video messages are viewed sequentially or simultaneously.

5. Fantasy Video TV studios and the Fantasy Video Broadcasting Station

In order to have something for the Fantasy Video TVs to pick up, it is necessary to have companies or groups that specialize in broadcasting Fantasy Video TV programs. The equipment they will need to do this consists of a Fantasy Video TV Broadcasting Station, as shown in Fig 10.

There are two methods for doing this, live, and prerecorded. Some improv comedy groups will want to perform live. In this case, all that is really needed is a Fantasy Video Sender, usually with multiple Image Acquisition Devices, for the one or more actors to use.

However, in most cases, the shows will be prerecorded and then edited. A Broadcasting Station thus needs two additional components that will typically be intermixed, a Means for Recording or a Recording System, and a Means for Editing or an Editing System.

The Recording System, consisting typically of a Means for Recording 1010 plus a Storage System 1020, is quite straightforward and is a component that is also used in the Fantasy E-mail Sender, etc. The Means for Recording is typically a simple routine that accepts the stream of information coming from the Imaging System and streams it into a file, instead of allowing it to be sent over the distribution channel.

The Storage System typically consists of a hard disk on which a new file is opened. In some cases, however, it might be a RAMdisk or a buffer in memory.

The Means for Editing has been previously discussed as a Multitrack Editor. It consists of a software Editing System 1030 that can read and write different tracks off of the Storage System, a local Presentation System for the editing viewer to review those tracks upon, and a user interface for the editing viewer to control the Editing System with. The Editing System allows modification of the enhancements and editing of the essential information involved in a scene, including such things as inserting, deleting, cutting, pasting, overwriting, tuning, changing, splicing, etc.

The finished Fantasy Video TV/Movie is then made available to a distribution channel.

5A Fantasy Video TV studio records actors using a Fantasy Video imaging system.

A typical day for an actor in a Fantasy Video TV studio will go something like a day for a radio announcer. Various scripts will be given the actor, and he or she will read the scripts in a dramatic manner into an Imaging System hooked to a Recording System while moving his or her body and face. The system will probably use a teleprompter program to let the actor read his or her lines on the fly.

Because the actual appearance of the actor can be replaced, including the voice, one actor can hold down many parts. This allows salary savings for small studios.

5B Fantasy Video Television, retrofit model

A small set-top box houses dedicated special-purpose hardware that accomplishes the functions of the presentation construction subsystem. It accepts incoming information from broadcast radio waves on television channels or on AM or FM radio channels for input as a distribution channel, or input from the Fantasy Video VCR. The Video TV box builds a presentation of users plus an environment. An S-video or composite video cable, etc., coming from the Video TV box leads to the viewer's TV and provides a video signal plus optional sound. Several different presentations may be built simultaneously, reflecting input on several different Fantasy Video TV channels, and presented simultaneously, using a split screen or window-in-a-window, etc. This device is provided for viewers who already have a TV.

5C Fantasy Video TV, inboard model

It is possible to take the external set-top Fantasy Video TV box of 5B and design it into a TV so that it is an integral part of the TV. Then customers can buy a new device that accepts Fantasy Video information, constructs the presentation, and displays it on an integrated TV monitor, all in one box.

5D Fantasy Video VCR

It is possible to record a Fantasy Video session on the side of the viewer, instead of on the side of the user as for Fantasy Video Email. In this case, the embodiment is a Fantasy Video VCR. This is useful in recording Fantasy Video broadcast programs or Fantasy Videophone conversations for later viewing. The recording can be done to a computer file, a CD-ROM, a DVD disk, or a magnetic tape such as a videotape. Instead of recording the literal video and audio image, as is done currently, only the essential information required for reconstruction is recorded. Then, later on, when the viewer wants to play the recording again, the viewer's system uses its presentation system to construct a presentation based on a stream of information from the recording, instead of from a live-source stream. As with the other embodiments, the viewer may dynamically change the format of the presentation, including such things as the users' costumes, the voice fonts, the environments, the camera angles, the lighting, and the modification filters, etc., at will, before or during playback, using appropriate software commands. These commands may be input live by the viewer or may also be recorded and later edited as part of the recording.

5E Fantasy Video VCR TV

Some people will want to record Fantasy Video TV programs that come across the distribution channel. In this case, the Recording System including a Means for Recording and a Storage System are used on the viewer's side instead of on the user's side. The Recording System will typically be connected with a software "Y" connection in the stream that allows the viewer to view the presentation at the same time that the Recording System is also recording an identical copy of the stream. Then a Playback System is used to view the recorded stream later.

5F Fantasy Video VCR TV Telephone

The same setup can be used to record Fantasy Videophone conversations from a Fantasy Videophone Station. In this case, the Recording System should record not only the incoming stream but the outgoing stream as well. The Recording System will then have at least two tracks of information streams per conversation; these can later be shown as a split screen, on two separate presentation devices, etc.

5B. TV set-top box that uses the telephone or cable-TV network.

In this embodiment, the Fantasy Videophone is instantiated as a small special-purpose hardware box that sits on top of a television (TV). An outboard or integrated video camera connects with an internal frame-grabber to allow the set-top box to capture images of the user as fast as possible, preferably at video rates. The distribution channel is the telephone network, the cable-TV network, or the cellular phone network.

6. High-definition wall-mounted flat panel TV/Fantasy Videophone, linked by telephone to a dedicated communication service over Internet. Mounted in front of a breakfast table, eat breakfast with friend in Japan.

This embodiment uses a flat-panel monitor for the presentation device. The flat-panel screen may be mounted on the wall, stood up on a table, mounted on a door or on a refrigerator door, etc. The monitor may be a high-definition TV, a high-definition computer monitor, a regular TV monitor, or of any other dimensions, etc. The camera required for a two-way conversation will typically be mounted above the screen, to one side, or at the bottom of the screen. The system may be a dedicated Fantasy Videophone; or it may be combined with one or more of the functions of a TV, a telephone, a Fantasy Video TV, a computer, a fax machine, a VCR, a DVD or CD music player, a laserdisc movie player, etc. The system may accept information through one or more of the distribution channels of the telephone network, television broadcasting signals, radio waves, cellular phone signals, the Internet, dedicated cable, etc. A typical embodiment will have a telephone wire coming out the back for use in two-way Fantasy Videophone calls, along with a television-signal receiver for accepting one-way broadcast Fantasy Video TV signals, along with a keyboard, computer modem, and general-purpose CPU to support regular computer and Internet usage.

A key feature of this embodiment is that the camera will typically not be lined up directly in front of the face of the user. The user will be facing the screen directly, but the camera will typically be above the screen looking at an angle down at the user. The user will also be positioned close to the screen, causing barrel distortion in the unmodified input image of the user. A key component of this system is a filter that modifies the apparent camera angle of the scene for the viewer. This may be done using several methods. Two are presented here.

6A. Computer Graphics Method The imaging system acquires pose information from the user, such as the amount of eyelid opening, the height of the eyebrows, the angle of the jaw, etc. This information is sent to the presentation system. The presentation system uses computer graphics methods to generate the presentation. A computer-graphics costume model is negotiated and specified, along with an environment image or model. Then, the presentation system specifies the location and lens angle of a virtual camera, which can be located mathematically behind where the user's screen would be in the scene. The virtual camera angle can point straight out horizontally from the screen, and the distance can be moved back behind the screen to a comfortable distance so as to frame the user well. The virtual lens angle can also be adjusted to a comfortable zoom or wide-angle factor. The virtual camera can be positioned and adjusted automatically, interactively by the viewer, interactively by the user, by a third party or program, or a combination of these, etc. The virtual camera can be still, it can track the user, or it can move a proportionate amount between remaining still and tracking the user so that it moves e.g. halfway towards the user when the user moves. By having the virtual camera be positioned behind and a comfortable distance back from the screen, the presentation of the user appears at a comfortable distance and angle. The user appears to be talking face-to-face with the viewer, instead of being seen from above with distortion from being too close.

10. Wrist-mounted Fantasy Video TV and Fantasy Videophone.

Cellular wristwatch picture phone.

10A. In this embodiment, the presentation device is a small system that is mounted on the viewer's wrist, belt, collar, armband, or is otherwise worn as a piece of clothing or carried by the user. The preferred embodiment for this is a device that is combined with a wristwatch, which may also tell time and store telephone numbers and appointments. Embodiment 10A is a Fantasy Video Wrist TV. It has a receiver that can accept a stream of input from a distribution channel, along with a presentation system that shows the viewer output-presentations on a screen or by projecting directly into the eye.

The system may also incorporate the functionality of a regular cell-phone, TV, fax machine, computer, answering machine, etc.

10B embodiment 10A, except that the device is a two-way Fantasy Videophone. The wristwatch device also contains an imaging system, having a small built-in camera for image acquisition and hardware for image information representation. There is also a means for transmitting information over a distribution channel. This could be a cell-phone channel, a radio link, a phone jack, an acoustic coupler for a telephone, etc. In a manner similar to the Wall-mounted Fantasy Videophone, the Wrist-mounted Fantasy Videophone can use software components for virtual-camera repositioning and distortion elimination.

10C embodiment 10A, except the Fantasy Video TV system is the size of a cellular telephone or notebook and is carried by hand instead of being mounted on the viewer's wrist.

10D embodiment 10B, except the Fantasy Videophone system is the size of a cellular telephone or notebook and is carried by hand instead of being mounted on the viewer's wrist.

10E embodiment 10D, except the system is also a Personal Assistant computer, such as the Sharp Mobilon. The user can use the computer, or can make Fantasy Videophone calls from the same convenient machine. Since the Personal Assistant camera will be held close to the user's face, it will again be necessary to apply a virtual-camera-repositioning filter to avoid barrel distortion.

10F embodiment 10A, except that the presentation device for the Fantasy Video TV system has a pair of special eyeglasses that allow the user to see and hear images.

10G embodiment 10B, except that the presentation device for the Fantasy Videophone system has a pair of special eyeglasses that allow the user to see and hear images.

7. Internet sitcom program, single actor version, recorded (lay down tracks), and broadcast on Web.

8. Internet sitcom program, multiple actors, live real-time version

9. A News program with a canned announcer:

9A There are a number of embodiments for implementing a News program. The most straightforward is to use Fantasy Video Email and record a program of an announcer reading the news, then make this file available for play-on-demand by viewers.

9B The second embodiment is to motion-record a skilled announcer reading a series of typical bulletins, and use an editor to clip out stereotypical tiny motion sequences. These can then be made to join smoothly and placed in a finite-state machine. Then a junior announcer can read the news. The junior announcer can use his or her own appearance, but perform the change known as "overriding" to insert the movement patterns of the skilled announcer.

9C The system executes the finite-state machine automatically, using the virtual costume of the skilled announcer. The junior announcer only has to read the text.

9D is 9C, except the text is read automatically by a text-to-speech program. However, the announcer's movements still come from playing back movements that were originally recorded from a person.

9E is 9C, except a professional voice actor should be recorded to get a voice font. Then this voice font can be used to replace the tonal qualities of the actual announcer.

9F is 9D, except a professional voice actor should be recorded to get a voice font. Then this voice font can be used to replace the tonal qualities of the computer announcer.

11. Personal assistant computer Fantasy Videophone Station or Fantasy Video TV

An important form-factor for the invention is that of a "personal assistant" computer, sometimes called a palm-top. It is about the size of a checkbook and typically has a color screen and a small keyboard. Newer models have a color camera built in, along with modem hardware and a phone jack. Cellular phone connections will also become popular. A Fantasy Video TV implementation would consist of a Fantasy Video Receiver as shown in Fig 1C, where the color screen is the Presentation Device, the modem or cell-phone modem is the Means for Accepting Information from a Distribution Channel, and the computer itself supports software embodying the Presentation Construction Subsystem. The camera could also be used as an Image Acquisition Device, along with the built-in digitizer, and the computer again supporting software for an Image Information Representation Subsystem, along with the modem again, to form a Fantasy Video Sender. Then the two software programs could be merged to form a Fantasy Videophone Station. Like

any other Fantasy Videophone Station, this form factor could also support the extra software components for Fantasy Video E-mail, an Editing System, etc.

12. Cyber Bar dating club or amusement park

A multi-user Fantasy Videophone cyberspace may be used for the purposes of running a dating club or running an amusement park. Each customer has a Fantasy Videophone. A central third-party company manages the distribution channel and takes responsibility for creating interesting environments and costumes. Each ordinary customer user is also a viewer. There are some viewers who are not users, called "lurkers" or "spectators". Spectators may watch on a Fantasy Video TV, since they do not need the imaging system capabilities. There are some users, typically hired by the company, that have special viewing capabilities and are called "actors", who put on performances for the customers and the spectators. Users and viewers connect to a central server complex that is run by the company, typically using a star topology. The company sends environments and virtual costumes to their presentation systems. Users use the Fantasy Videophones to talk and interact with other users and with actors.

13. Sports arena multi-to-many broadcast.

A multi-user Fantasy Video TV cyberspace may be used for the purposes of running a sports arena. Actors called "players" and a referee are equipped with two-way Fantasy Videophones. The spectators are equipped with Fantasy Video TV sets, or they view the sports using the presentation system of their Fantasy Videophones. The players act out a sport by their movements. The spectators watch the sport. A central company provides support for the distribution channel, the environments, and the virtual costumes.

14. Recorded cyber programs distributed on CD-ROM.

It is not necessary to broadcast Fantasy Video TV programs over the Internet. A company can place hours of entertainment on a CD-ROM or tape and sell it to customers with Fantasy Video Receivers. It is necessary to have a Playback System built in to the Receiver to open the file on the CD-ROM or tape and send it as input to the Presentation System; this Playback System counts as part of the Distribution Channel in this case.

15. Fantasy Videophone Call Center for order taking, technical support, and instructional purposes.

15A in this embodiment, the Fantasy Videophone system is used for call center applications. The distribution channel has a central call-routing device that can accept multiple incoming calls. The call center also has a database, and a database management system. A call center department sets up a plurality of operators. Each on-duty operator is furnished with a

Fantasy Videophone Call Center Terminal. The Terminal is a regular two-way Fantasy Videophone with additional software to display help screens, display menus, and take orders. A customer also has a two-way Fantasy Videophone. The customer places a call in to the call center. The central call-routing device finds an operator who is available, and connects a two-way Fantasy Videophone call between the customer and the operator. The operator can provide technical support, can describe products and take orders, can provide educational instruction, can provide counseling, or can provide other forms of support, etc. The Terminal provides an interactive display for the operator to read from and write or dictate into, in addition to the Fantasy Videophone functions. If the customer requires a supervisor, the supervisor can be patched in into a three-way etc. conversation using an additional Fantasy Videophone. If the local telephone network has caller ID, this information can be displayed on the screen for the operator and inserted into the order automatically. The database management system can also call up other significant information that is already known by the call center about the customer. This information can be displayed symbolically by changing the presentation of the customer user and the customer's environment. For instance, if the customer is known to have already bought product from the call center's company, or is known to drive a red Porsche(tm), then these items can be inserted in miniature or full size etc. into the environment's presentation.

The Fantasy Videophone Call Center can be set up so that the operators are always wearing a standard company virtual costume. This can be a uniform substituted in for the operator's clothes; it can be a standard fashion model that each operator uses for presentation; it can be a company mascot; it can be a color that the operator's presentation turns, or objects placed on the operator's presentation's head, etc. Different costumes can be used to differentiate order-takers from technicians, etc. A separate name card or plaque containing the operator's name or ID number can be presented to the customer.

15B The same as 15A, except the distribution channel uses the Internet to accept incoming calls. Calling customers use a Fantasy Videophone attached to the Internet, use the web, or use e-mail to submit orders.

15C The same as 15A, except the distribution channel uses computer telephony over the Internet to accept incoming calls. Customers use a Fantasy Videophone attached to the telephone network or to the Internet to place calls.

15D The same as 15A, 15B, and 15C combined.

15E The same as 15A, except the operators are furnished with standard telephones for incoming calls, and the Fantasy Videophone is only outgoing. The customer may have a Fantasy Videophone or a Fantasy Video TV.

16. Call service for sexual entertainment.

A) Single performer, one-way A sexual performer can run a videophone-based business from home or a nearby warehouse. Because the Fantasy Video system can replace the possibly-clothed performer with arbitrary unclothed movie stars of fantastic proportions, and can replace the background with arbitrary scenic sets, this invention enables amateurs of limited means to enhance the economy.

a) multiple performers, one-way. Many different users in different locales can get together to create a performance. A central computer can assemble their presentations together into the same scene.

b) Broadcast large audience The business does not have to be run for single callers at a time, but can be a performance for multiple viewers.

c) Interactive two-way. A Fantasy Videophone Station allows a user to place him or herself actually in the scene along with other actors, and then watch the performance from a convenient bird's-eye or over-the-shoulder viewpoint.

17. Live Fantasy actor commenting on top of a multimedia performance

A commentator can be composited in on top of a television or multimedia performance. The commentator can be transmitted using a Fantasy Video stream, while the other channel is transmitted over the same or a different stream of information and then composited by the Presentation System.

18. Multiple point-to-point conference calls for business.

It is straightforward to set up conference calls amongst multiple Fantasy Videophone Stations. One simply has to connect them in a Mesh topology 0720 or one of the other multi-caller topologies discussed in Fig 7.

19. Fantasy Videophone broadcasting system with editor.

19. Laser rangefinder system with editor that sends to a remote sculptor tool.

It is not necessary to acquire moving images. A portrait-carving company can have a computer-controlled machine that carves sculptures as a Presentation Device, and a laser-based rangefinding scanner machine that acquires spatial occupancy data as an Image Acquisition Device. Then the company can acquire the spatial image of a user, abstract it from its environment, ship its essential information across a distribution channel, possibly perform enhancements on the image, and present the enhanced image as a sculpture.

21. Another embodiment using a rangefinder.

The imaging system can use a rangefinder as input in most of the previous embodiments instead of using a vision-based imaging system to acquire spatial information. If the user is being completely

replaced, or if a virtual costume of the user's body is on file, then there is no real need for the actual image information in the system. The imaging system uses the range image of the scene to abstract the pose of the user and abstract essential information from that. The rest of the system proceeds as before in the various embodiments.

In this embodiment, the imaging system uses a rangefinder such as a laser rangefinder to acquire a physical depth map of the user and the environment. Then the information is sent over a distribution channel to a presentation system, which uses a sculpting tool or hologram system to make a physical presentation of the scene. An example-sculpting tool is a system that uses a laser to carve out wax. The user and/or the environment may be changed by the system, edited interactively, or removed, in the imaging system, or in the presentation system, etc. In this manner, a physical portrait of the changed user and/or environment is created remotely by the system.

22. Fantasy Video System using a perspective-morph based method for representing and presenting information

In another embodiment, the imaging system does not directly perceive and abstract a set of 3D polygonal features as the underlying representation of the system. Instead, when performing information representation, the system uses a set of 2D "eigenfaces" for each object to be recognized and worked with, that together contain enough information to be able to represent the appearance of the object at any orientation. Position and size are typically normalized out. The set can be dense, in which case it is usually represented as a 4D space called a "light field"; or it can be a sparse graph of eigenfaces partitioning the space. An image of the object at a particular orientation can be indexed directly by the 4D coordinates in the dense representation, or by a small set of eigenvalue coordinates together with the identities of the nearest eigenfaces in the sparse representation. Recognition of the object and its orientation can be performed conceptually by convolving the image with the representation; in practice, this is done by using a neural network.

The imaging system is trained on all objects to be recognized. Then, during operation, recognition of the object and its location is performed, to abstract the image of the object into an essential representation having the coordinates. These are then shipped down the distribution channel to the presentation system.

The presentation system can use these coordinates to re-create the appearance of the object, or to create the appearance of a replacement costume object, by using a similar light-field space or eigenface set. The dense space or sparse set can represent the appearance of the object itself, or of a replacement object. In the case of the light-field space, the coordinates index into the 4D field and a 2D image is yielded directly. For nonintegral coordinates, some interpolating may have to be done between the nearest neighbors at the integral grid corners in the space. For the sparse eigenface approach, the nearest eigenfaces are indexed, and a perspective morph is performed, using the eigenvalues as mixing coordinates,

to come up with the resulting appearance. Perspective morphing requires a list of corresponding feature points, but works with 2D input images to give a 2D photographic output image corresponding to the proportional rotation of the object in 3D. For more details, see "View Morphing", Seitz and Dyer, Siggraph 1996.

The imaging system does not have to work with 3D information. In another embodiment, the imaging system can abstract 2D image-coordinate features, such as edge outlines, and send these down the distribution channel to the presentation system. Then the presentation system can create a 2D presentation. This does not have to be photorealistic; it can look like an oil painting, a line drawing, a cartoon, etc.

As the invention is meant to be an improvement on current videophones or televisions, its main method of usage should be intuitively obvious. The user positions him or herself within range of the user's Image Acquisition Device(s) (for example, a camera with a microphone) and speaks while making bodily and facial gestures as desired. The viewer watches a presentation of the user by positioning him or herself within range of the viewer's Presentation Device (for example, a TV or computer screen with a speaker). In this manner, the user can use his or her Fantasy Video Sender, and the viewer can use his or her Fantasy Video Receiver. If they both have Fantasy Videophone Stations, they can both speak to each other.

Since one of the key features of this invention is the fact that either the user, or the viewer, or both, can choose to enhance the presentation by changing its characteristics, both the Fantasy Video Sender and the Fantasy Video Receiver will typically have controls or menus on them that allow specification of the particular changes that are desired. For instance, the user might choose to swap out the environment with an image of a beautiful beach, complete with background sound effects, while the viewer might choose to replace the body of the user with that of a giant shrimp.

23. Embodiment

A simple embodiment of the present invention runs on a Pentium™-class PC computer running Windows 95™. The PC should be equipped with an ATI All-In-Wonder™ Pro video digitizer card or equivalent, along with installed drivers. The computer monitor display should be set to "True Color (24 bits)". A color camcorder, such as the Minolta 8100, or similar video source should be used as video input. The video source should be set to a fixed focus option (as opposed to autotracking autofocus); the aperture should also be fixed; and, if possible, the white balance should be fixed. Auto-tracking automatic white balances are a significant source of problems for the system. The camera should be placed above the computer monitor screen so that it can image the user, who is to sit in front of the computer screen.

For an intialization period of preferably about five seconds, the system displays a message. During this time, the user must carefully vacate his or her seat, without disturbing the spatial arrangement of environmental objects (including the seat) within range of the camera, so that the camera can no longer see the user. After the initialization period, the user should resume his or her seat, again without disturbing

the location of any environmental objects. At this point the location of the user is now being tracked by the system.

Hitting a first key on the computer will delete the presentation of the background environment and substitute in a 2D picture of a luxurious hall. The user can move about at will, and the system displays the user in front of the Fantasy environment. This experiment demonstrates replacement of the environment.

Hitting a second key on the computer switches the environment to a 3D virtual office. The image of the user is placed in a flat plane in front of the back wall and chair, but behind the desk. This demonstrates that the presentation of the user can be placed inside a 3D virtual scene, and not simply overlaid on top of its image.

Hitting the a third key on the computer switches the actual environment back in, but replaces the user with a 3D-graphics humanoid avatar costume of a robot. The avatar tracks the XYZ position of the actual user in the scene. If the user moves left or right, the avatar stand-in moves stage-left or stage-right on the screen. If the user moves up or down, the avatar moves up or down. If the user moves forwards or backwards relative to the camera, the avatar moves forwards or backwards.

If the user turns his or her head to the left or right (yaw), the avatar presentation turns its head a similar amount. The example embodiment currently does not yet support other rotations, such as tilting the head sideways (roll) or nodding the head back and forth (pitch). The current embodiment also does not support tracking arms, hands, or legs, and the trunk is assumed to be vertical and aligned facing the camera throughout.

The system works by patching into the Microsoft video input stream. This sends a stream of video frames to the program, for it to work with.

The first step that the system takes is to reduce the data from a 320x240 array to an 80x60 array by averaging pixels over a 2x2 grid twice in a row. This could equivalently be done by averaging over a 4x4 grid. The resulting video image is a low-pass spatial filtering of the original image, which acts to reduce both salt-and-pepper noise and the amount of data required to be worked with. A three-frame temporal filter is used as well to reduce jitter.

When the user vacates their seat, the system takes a snapshot of the empty scene. This is reduced by a factor of 4x4 and stored in a buffer.

When the user resumes their seat, the system continually processes the scene by taking snapshots, reducing them, and comparing them against the saved empty environment buffer.

Comparison is currently done using a texture metric along with a brightness metric. This has the advantage of being usable by both color and black & white systems. A stable system should consider using a hue metric. However, most low-grade systems have automatic auto-white-balance functions. With the

white background walls commonly found in America, this function results in significantly different settings between identical environmental scenes lacking a user and having a user. A white wall without a user could have an orange cast, whereas with a user the exact same wall could have a green cast. For this reason, it is safer with current-technology cameras to use thresholding routines that do not depend on hue.

A texture metric is computed by taking a 2x2 window and adding the two pixels on the left diagonal while subtracting the pixels on the right diagonal. Any pixels that are the same as the environment will tend to have texture metrics that are identical with those of the empty environment image, which is computed once and cached in a buffer, whereas any pixels that are user pixels will tend to have different metrics. Any pixels that have wildly differing brightnesses are classified as "user" as well.

It is assumed for this simple system that there is only one user, and that the user has only one head and is not raising his or her hand. It is also assumed that the user is seated generally in front of the computer. A salt-and-pepper filter is run on the classification image to flip outlying pixels that have only zero or one neighbors of the same color. Then the system encodes the horizontal array scanlines of the image into scan runlengths. Only the longest scan runlength is kept for each scanline, and the rest are considered garbage and deleted. The scan lines are further filtered in the Y direction for consistency; any scanline that is significantly different from both the one above it and the one below it is considered anomalous and is adjusted to the average of its neighbors' endpoints. The result is a classification image consisting of a silhouette of the user's head and shoulders. This is now ready for blob analysis.

The picture Y axis measure of the top of the user's head can be picked off by seeing at what point the scan-lines start coming down from the top of the picture. The shoulders can be found by moving downward, keeping a moving average of the runlength widths, and finding at which point the widths seriously increase and then stay increased. Everything below this is torso and shoulders; everything above this is head and neck. The picture X axis measure of the head can be found by averaging the centers of the head/neck runlengths, and similarly for the torso. The height of the shoulders is determined by where the shoulder cutoff was found. The total area of the head determines a rough estimate as to the Z depth from the camera of the user.

The central horizontal axis of the head determines a place to start searching for the eyes, and thereafter the eyes are seeded by their previous offsets from the center of the head. The eye sockets are found by convolving the low-pass brightness image with a number of sizes of generic eye socket images. The two best consistent results are chosen. The position of the eye sockets relative to the head blob tells the orientation of the face. The irises can be found at a higher resolution in the eye sockets by searching horizontally for a large scan of a dark color surrounded on both sides by a large scan of white; the two iris blobs can be found from this. The corners of the eye and the eyelids can be found from finding the edges in the image in the eye socket region. When there is no iris, the lids are closed. The center of the irises relative to the corners of the eyes and the orientation of the head determine the gaze directions.

The mouth can be found from its expected position and because the lips are dark. The top of the central portion of the upper lip, and the bottom of the central portion of the lower lip, determine the lip opening. If both upper and lower white teeth can be seen inside the mouth, they determine the jaw opening; otherwise, the approximate jaw opening is determined from finding the shadow under the chin to determine the point of the chin. The corners of the mouth are pulled out of the edge image by growing the strong edge of the underside of the upper lip sideways. It is currently assumed that both mouth corners are visible; it would be necessary to compare the mouth image against a standard head solid model in order to catch horizon effects if this assumption were disregarded. The mouth literal texture must be inside the head blob and is found by taking the max and min in Y of the top and bottom of the lips and the mouth corners, along with the max and min in X of the mouth corners. A more complex algorithm would grow dark blobs to find the lips and take everything between the top and bottom lips. The eye literal textures are found by taking everything inside a rectangular eye socket region.

Thus, we have extracted the following essential information from the image:

Literal texture image of the environment

Classification binary image of user vs. environment, resulting in literal texture image of the user

Head (x, y, z) and head orientation, expressed in a 4x4 matrix (with implicit last row)

Height of shoulders

Picture X coordinate of the torso

Eye location and gaze orientation, expressed in two 4x4 matrices

Corners of eyes and eyelid opening

Jaw opening

Lip opening, lips' top and bottom, and the positions of the corners of the mouth

Literal textures for the eyes and the mouth

The literal texture image of the environment is sent across upon initialization. The literal texture image of the user is sent across when it is desired to present a literal image of the user, perhaps with an environmental replacement; this selection can be made by the user or by the viewer, in which case the Fantasy Video Receiver sends a selection message to the Fantasy Video Sender. The literal images of the eyes and mouth are sent across under certain enhancements. The classification run-length binary image is not used for transmission nor by the presentation system in this system. Otherwise, all of the other information is packaged up in a structure, along with a header code, and sent to the Fantasy Video Receiver on the remote computer by Winsock sockets. It could also be sent as a plug-in application using the

Microsoft NetMeeting transport system. In the present design, it is necessary to establish these distribution channels by hand.

On the Fantasy Video Receiver side, the essential information is taken from its packaging structure and used to help create a presentation. The environmental literal texture image is received and cached upon initialization. The Receiver also opens up some local formatting information files, including

- a photograph of a luxurious hall
- a 2D pictorial image of a dungeon cyberbar
- a 3D model of a cyberspace office, including a desk and chair in a room with props
- a 3D model of a humanoid robot with head, body, arms, legs, eyes and a jaw

The actions of the Fantasy Video Receiver depend upon which enhancements are requested.

If a simple environmental replacement is requested, the environment behind the user is replaced with either the photograph of the luxurious hall, or the picture of the dungeon cyberbar, as specified. The literal texture image of the user is overlaid on top of this. The empty parts in the literal texture are run-length encoded so that they don't take up any space; this reduces the size of the image by almost a factor of 3:1, depending on where the user is sitting. The presentation construction subsystem builds a pictorial image in a buffer in memory, and then displays it to the screen. In this case, the buffer is built in a 2D manner by copying pixels from either the chosen environment or from the literal texture image of the user.

If a 3D environmental replacement is requested, the literal texture of the user is mapped onto a large otherwise-transparent rectangle, which is positioned in the space where a user would normally be in the cyberspace environment, i.e. behind the office desk sitting in the chair. The rest of the office is drawn using Direct3D, Microsoft's solid modeling system. It is necessary to specify a virtual camera position. The lighting can be left on the default, or virtual lights can be specified explicitly. With this enhancement, it is not necessary to use the other essential information, such as the xyz position of the user, since this will be reflected in the literal texture image. However, it would be easy to connect the position or velocity of the picture of the user with the X or Y position of the user relative to the camera.

If a user replacement costume is requested, the system can use the avatar and display it in an appropriate pose using the cached literal environment picture as a backdrop. The model is created using Direct3D again. It is posed by using the essential information variables for the positions and orientations of various parts, along with a "wiring" structure that instructs the program as to which variables to change in order to move which body parts. For example, the position and orientation of the left eye is in a particular matrix in the 3D model, so the "wiring" has an association between "left eye" and a pointer to these variables. Then, when the essential information structure comes in with the "left eye" information labeled, this information is copied from the structure into the contents of the location pointed to in order to effect

posing of the left eye. The eyes, the head, the body, and the jaw can all be positioned appropriately. In this manner, the replacement costume is posed in the scene using the essential information, corresponding to the user's pose.

Besides merely replacing the presentation of the user, it is possible to replace the presentation of the environment at the same time. The avatar can be presented in front of either of the 2D environments. Or, the avatar can be presented in the midst of the 3D office cyberspace environment, again using Direct3D to construct the images of both.

Finally, everything except the user's eyes and mouth can be replaced. Both eyes and the mouth are sent across the distribution channel as literal textures, and used on the avatar's head instead of the solid models for eyes and mouth.

24. Robotics application of the present invention

Once the Fantasy Video Receivers become prevalent, people will want to create centralized robotic artificial-intelligence programs that stand in for users, for use as robotic actors in games and advertising. In one embodiment, a Fantasy-Video Robot program should run alongside an Image Information Representation Subsystem. The goal is for the Fantasy-Video Robot program to generate the same types of essential information in an outgoing stream as a person would generate.

One method for doing this is to have a Fantasy Video Recorder record a number of basic units of action from a real user, such as "shrug shoulders", "wave right hand", "wave left hand", "nod head enthusiastically", etc. into a Storage System such as a computer disk. Then use a Fantasy Video Editor to cut these recorded motions up into separate tracks, each of a few seconds' duration, and label them accordingly. Then build a software state-machine that is wired to these tracks on the disk. The back end of the state machine is an artificial intelligence having a number of state variables, and a number of rules for changing states based on other states. For instance, one state variable could be "amount of excitement", on an integer scale of 0 to 10; another state variable could be "amount of hunger" on an integer scale of 0 to 10. There could be a software rule in the system that says, "If the amount of hunger state is greater than 5, then add 1 to the amount of excitement state until it reaches 10." A software clock is useful to run the state machine; every second or every 1/30th of a second, each of the rules gets examined and executed if appropriate.

The middle part of the state machine is a state that is labeled "current behavior". It stores an integer code that determines what behavior is going on. For instance, 1 could be "shrug shoulders", 2 could be "wave right hand", etc.

The front end of the state machine is a software routine called "wiring" that takes care of executing behaviors. It has an integer-indexed array that stores the file names of the behavior streams on the disk. For instance, array entry number 1 stores a string "ShrugShoulderFile.FV", array entry number 2

stores a string "WaveRightHandFile.FV", etc., where the .FV extension indicates a Fantasy Video recorded stream of essential information. The front-end wiring routine observes the "current behavior" state and compares it against an internally-stored static variable to see whether it has changed or not. When it changes, for instance the "current behavior" goes from 2 to 1, the wiring routine notices and executes behavior number 1, "shrug shoulders". Behavior number 1 is executed by opening the file that the wiring points to, reading its contents, and streaming those contents to the output stream of the Fantasy-Video Robot. This will typically be connected to a distribution channel, although it could be connected to a Fantasy Video Recorder for further use.

The back-end state machine can be of arbitrary complexity. The state machine connects to the outside world by having rules that change the "current behavior" state variable based on various other states. For instance, if one makes an "if amount of excitement is over 5, then current behavior is 2='wave right hand'", then the robot would get excited and start waving its right hand whenever it starts getting hungry. If a rule such as "if current behavior is 2='wave right hand', then set current behavior to 3='wave left hand'" is implemented, then the robot can execute a series of actions in a particular order. These again can be made as complex as desired.

Typically one of the actions will be standing still or fidgeting back and forth in a neutral pose. Then other actions are recorded that start from this neutral pose, do something, and return to the neutral pose. In this way, different actions can follow each other in different orders, but there is still no break in the continuity.

The resulting performance is sent across a distribution channel and viewed by interested viewers on their Fantasy Video Receivers.

A second embodiment uses speech generated by an artificial-intelligence robot. Since speech, based on sound images, is simply another modality for output, the same implementation architecture can handle speech output. In the straightforward case, the speech consists of recorded paragraphs giving a particular advertising spiel or describing a particular function for a help desk. Then the robot can choose to play back any particular paragraph based on its current state in its state machine. In more complex systems, the robot includes a natural-language generator that creates appropriate paragraphs composed of words on the fly. Then the front-end wiring consists of a dictionary of pronounced words that are stored on the Storage System (disk) and indexed in an array. The natural-language generator runs based on what the robot is interested in saying, which again is all driven by the state machine. Note that if the system uses a Fantasy Video distribution channel for output, it is not necessary to generate the speech itself, but simply the essential information describing the speech.

It is also possible to build an embodiment that comprises a Two-Way Conversing Fantasy-Video Robot. Such a robot needs to accept information from the viewer(s) in order to carry on a conversation. The information can be in the form of mouse clicks on Web buttons, or pointing at different

desired items, or typed text, or speech input, or other forms of input such as eye blinks or head nods. The input will typically come over a distribution channel from a user acting as a viewer of the Robot.

Such inputs as mouse clicks are straightforward; the system compares the 2D screen-space point of the mouse click against the 2D projections of 3D objects in the scene space, determines the frontmost candidate for clicking, and sends a "mouse clicked" message to the selected object, which picks up the message and calls an appropriate handling routine. If this is integrated with the Robot, it could trigger such behaviors as having the robot actor turn and wave his hands at the selected object, and start to explain it. Handling the user pointing at different objects requires a routine that solves for the geometry of determining which scene object on the screen the user is pointing to. Since the essential information describes the joint angles and configuration of the user, the system must take into account the Presentation Device's position in the actual world in front of the user acting as a viewer, cast a ray in Euclidean space from the end of the user's finger to the Presentation Device's presentation, and proceed as if this were a mouse click.

Typed text is straightforward; if the text consists simply of keywords, then these words are entered in to the artificial-intelligence state machine using a "the keyword that the user typed was" state variable. If the text consists of sentences, then a software dictionary is connected to a syntactic/semantic sentence and paragraph parsing program, which is connected to a natural-language understanding program, which is connected to a user-intention understanding program. The robot reacts to the user's intentions and the semantic content of such intentions. For instance, if the user says "I'd like to buy the car", the intention is a buying action, and the semantic content is the particular car currently being shown on the screen. If the user next says, "I'd like to buy the pony", the artificial intelligence uses the same routines to respond to this intention, but changes a semantic-content variable passed in to the routine describing how it should react to such a user's stated intention.

A simple implementation of an interactive conversing program for amusement purposes can be built by adapting an ELIZA program to generate textual output (based on the user's textual input), which is then fed to the speech-information generator.

Speech input can be treated as a special case of typed text, in which the input is noisy and the actual words used are not understood completely clearly but may be ambiguous. The system needs a speech recognition system that takes as input the essential information from the user's speech, and gives as output a series of words for input to the software dictionary. The artificial intelligence behind the robot actor must then take into consideration the fact that it might not have understood the user's utterance in an accurate manner.

Eye blinks and head nods are semantic messages that can get passed in to the top of the artificial intelligence system. The system must have a hand-coded message handler that picks the message up and reacts to it in an appropriate manner, which is usually situation-dependent. This is exactly the same method

used today to handle keystroke or mouse-click messages by a series of different windows on a screen, where a mouse click could signal an application that the user wants to start typing inside a window, or it could signal a window that the user wants it to close itself—two very different actions handled appropriately by one system using messages.

The artificial intelligence should maintain a pool of the concepts that it wants to communicate, and it should be predicting ahead of the user what concepts the user might want to communicate. The artificial intelligence must select the next concept to be communicated, and must continue trying if it fails to get its point across.

In this manner, an artificial intelligence with a virtual robot actor can hold a two-way conversation over a Fantasy Video Circuit. The intelligence can take as input various types of information coming from the user, including Fantasy-Video essential information, speech input (which may be sent over a parallel distribution channel or may be encoded as essential information over the main distribution channel), visual gestures, text, mouse clicks, etc.

25. Fantasy Video E-mail Sender

One embodiment is a Fantasy Video E-mail Sender. Mailing under the Unix operating system is done in two main steps: (1) Create a file that is the desired recording of the stream of information to be sent; (2) Send this file out as e-mail. Obviously, the user must have already invoked the Fantasy E-mail Sender, and must already have specified the e-mail address (typically including the domain name) of the target viewer, along with a title for the e-mail. This can be accomplished with a screen prompt if necessary.

The first step is done by using a Fantasy Video Recorder, being a tiny Means for Recording subroutine along with a Storage System that is a computer hard drive, which is hooked to the back of software embodying an Image Information Representation Subsystem as part of an Imaging System. The Imaging System acquires images of the user, abstracts their essential information, represents this information, and passes it out in a stream to the Means for Recording. The Means for Recording first opens an empty temporary file on disk, and then simply copies each buffer in the stream onto the end of this file as buffers are handed to it. Then the Means for Recording closes the file when the message is ended. It should also have a safety function that closes the file and pops up an error message for the user if the computer hard drive runs out of space. More complex systems will allow one temporary file to be spread across multiple hard drives in case of unusually large e-mails or unusually small hard drives. In any case, the result of the operation is a recorded temporary file on disk that represents the information.

The second step is done by calling a low-level operating-system "mail" command on the file. This is done under Unix by building a string internally that invokes the "mail" function with arguments of the target's email address, including the target's host-computer domain name, and the name of the temporary file. A "system" command is invoked on this built string, which forks and executes a separate mail

process. Then the temporary file is "unlinked", which causes the temporary file to be deleted but only after the forked mail process is finished with it. The tiny routine that accomplishes these steps constitutes the Fantasy E-mail Channel Sending Subsystem. In some cases, such as sending pre-composed mail out through a Web browser such as Netscape, this functionality will be handled by a separate program.

The system must build in to the stream enough information for a Fantasy Video E-mail Reader to be able to play the mail. Typically, at the beginning of the stream recorded in the mail file, the system will include the negotiation information and formatting information to be able to replicate what the user requests for an enhanced presentation. The information is sent in chunks. The negotiation information will include a one-byte chunk code defining a negotiation-information chunk; the user's request for enhancements; references to well-known enhancement routines and multimedia properties; and references to included formatting information. For instance, the negotiation information could request a "BEACH" environment enhancement, a "SUNTANNED" filter for user, and a "TOWEL OVER SHOULDER" augmentation. The formatting information will probably include a single picture that is the abstracted user's environment without the user; a solid model or morph-space model of the user; a voice font of the user, if sound is being encoded; and literal copies of enhancement routines (e.g., Java source code) and multimedia properties (e.g., a photograph of a beach, and a solid-model of a towel) from the user's side that are not well-known. Based upon negotiation priorities, the viewer is free to change these into enhancements chosen by the viewer, or not, as the case may be. The formatting information is bulky and may be not included in some cases; in this case, the viewer's Presentation System will substitute in a standard environment, and standard models including graphics and voice, for creating the user's presentation.

The stream of dynamic essential information will also have chunk codes, defining encoded visual frames, encoded sound segments, or unencoded visual frames, sound, multimedia, or Web properties. For instance, it is possible to have a Fantasy Video sending essential information describing the visual picture while a separate track transmits sound that is unencoded or compressed in a normal fashion. Or, it is possible to have a Fantasy Video sending essential information describing the speech of the user while a separate track transmits movie frames that are unencoded or simply compressed in a normal fashion. It is also possible to send Web pages down along a separate track, by using chunk codes that define which track a chunk of information belongs to. Different chunks will have different codes that tell how they should be interpreted; essential information describing the facial and bodily pose of a user will therefore be appropriately treated in a different fashion than essential information describing the literal image of the user without a surrounding environment.

When e-mail is sent over a distribution channel as a file rather than as a stream, the receiving system on the viewer's side must be slightly different. Rather than having the distribution channel stream the essential information directly in to the Presentation System, a low-level function picks up the file and copies it into an email buffer in a Storage System in the viewer's Fantasy Video E-mail Receiving System.

This low-level function is typically already part of the viewer's computer's operating system.

Once the e-mail is held in a file in a local Storage System, an extremely simple routine opens the file up and streams it in by copying it, buffer by buffer, as input into the user's Presentation System. In the Unix operating system, this functionality is already automatically supported. In other operating systems, it could require building a "pipe" between a reading program and the Presentation System, or simply giving the Presentation System the capability of reading from a file. This is the Means for Playing Back a Message.

More complex systems will allow fast-forwarding, rewinding, and jumping to any particular time in the Presentation. This requires random access into the chunks supporting the tracks in the stream of information.

The Means for Playing Back a Message and the Presentation System taken together constitute a Fantasy-Video-E-mail Playing System. Such as system will be useful as a plug-in to existing Web browsers and mail readers in order to read single messages. However, it is also possible to build a stand-alone system that has added capabilities, for use by people having a Fantasy Videophone Station that does not have a complex operating system of its own that already handles reading e-mail. In this case, a message-selection system should display a linear or hierarchical menu of all of the mail in the viewer's mailbox buffer, and should handle such things as archiving, forwarding, replying-to, saving-as, and deleting mail. These functions are standard and are well-known by people who write mail handling systems. The menu display goes through all of the relevant mail files, gathers their titles, and displays the titles as part of a menu. Then the viewer chooses a particular mail file for viewing. Its contents are identified as holding a Fantasy Video stream by the mail viewer by means of a code in the e-mail or its header. Then Fantasy-Video-E-mail Playing System streams the contents of the chosen e-mail into the Presentation System by copying them in buffer by buffer, or by pointing the Presentation System at the location of the e-mail file or its contents and instructing it to start reading for itself. In this way a more complex Fantasy-Video-E-mail Playing System supports a user interface that chooses from multiple messages.

The Presentation System derives formatting information from chunks included in the stream, or from formatting information built in to the Presentation System, or from a local Library, or from a Library of Formatting Information that is part of the distribution channel or is otherwise available to the Presentation System. If required formatting information is not available, the Presentation System has to supply default formatting information, as usual. The viewer typically has the option of overriding the user's choice, and selecting enhancements for the presentation that are desired by the viewer. In this way, the Presentation System deals with presenting the essential information included in the e-mail message.

While the above provides a full and complete disclosure of the preferred embodiments of this invention, equivalents may be employed without departing from the true spirit and scope of the invention.

Therefore the above description and illustrations should not be construed as limiting the scope of the invention that is defined by the appended claims.

APPENDIX 1: "POSER™ 3" CONTROL PARAMETERS

The popular program "Poser™ 3" by MetaCreations™ allows a person to interactively pose a computer-generated presentation of a human or animal body. It demonstrates the current art and shows that using essential information to generate presentations of bodies, hands, faces, and environmental props is well within the state of the art. It illustrates a good starting point for designing a protocol for the content of the essential information.

Body Parts: Abdomen, Chest, Head, Hip, Left Collar, Left Foot, Left Forearm, Left Hand, Left Shin, Left Shoulder, Left Thigh, Neck, Right Collar, Right Foot, Right Forearm, Right Hand, Right Shin, Right Shoulder, Right Thigh

Shape Controls for each Body Part: Taper, Scale, X Scale, Y Scale, Z Scale

Movement Controls for each Body Part: Twist, Side-Side, Bend

Global Positioning for the Entire Body: X Tran, Y Tran, Z Tran
(global orientation is handled by the hip rotation relative to the world axis)

Articulated Hand allows control of all finger joints, plus globals: Grasp, Thumb, Spread

Basic Hand Poses are handled by a code running from 0-19; others are provided

Mouth Parameters for Articulated Face: OpenLips, Smile, Frown, Mouth-O, Mouth-F, Mouth-M, Tongue-T, Tongue-L

Eyebrow Parameters for Articulated Face: Left/Right Brow Down, Left/Right Brow Up, Left/Right Worry, Blink

Eye Poses for Articulated Face: (Left and Right separate) Up-Down, Side-Side, X Tran, Y Tran, Z Tran

Preset Faces, including Phoneme Faces: 7 Basic poses plus 17 Phoneme poses for entire face

Hair Augmentation: 11 wigs to add to figure

APPENDIX 2: FACS CONTROL PARAMETERS

This appendix lists the so-called "Action Units" (AUs) specified by Ekman and Friesen's Facial Action Coding System("FACS"). These form a good base for specifying a framework for essential information for facial expressions, and are used by most research systems in laboratories today. Some codes did not have muscles associated, and some numbers (e.g, 3, 40) are unassigned to codes.

AU 1:	Inner Brow Raiser	Frontalis, Pars Medialis
AU 2:	Outer Brow Raiser	Frontalis, Pars Lateralis
AU 4:	Brow Lowerer	Depressor Glaubellae; Depressor Supercilli; Corrugator
AU 5:	Upper Lid Raiser	Levator Palpebrae Superioris
AU 6:	Cheek Raiser	Orbicularis Oculi, Pars Orbitalis
AU 7:	Lid Tightener	Orbicularis Oculi, Pars Palebralis
AU 8:	Lips Toward Each Other	Orbicularis Oris
AU 9:	Nose Wrinker	Levator Labii Superioris, Alaeque Nasi
AU 10:	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
AU 11:	Naolabial Furrow Deepener	Zygomatic Minor
AU 12:	Lip Corner Puller	Zygomatic Major
AU 13:	Cheek Puffer	Caninus
AU 14:	Dimpler	Buccinnator
AU 15:	Lip Corner Depressor	Triangularis
AU 16:	Lower Lip Depressor	Depressor Labii
AU 17:	Chin Raiser	Mentalis
AU 18:	Lip Puckerer	Incisivii Labii Superioris; Incisivii Labii Inferioris
AU 19:	Tongue Out	
AU 20:	Lip Stretcher	Risorius
AU 21:	Neck Tightener	
AU 22:	Lip Funneler	Orbicularis Oris
AU 23:	Lip Tightener	Orbicularis Oris
AU 24:	Lip Pressor	Orbicularis Oris
AU 25:	Lip Part	Depressor Labii, or Relaxation of Mentalis or Orbicularis Oris
AU 26:	Jaw Drop	Masetter; Temporal and Internal Pterygoid Relaxed
AU 27:	Mouth Stretch	Pterygoids; Digastric
AU 28:	Lip Suck	Orbicularis Oris
AU 29:	Jaw Thrust	
AU 30:	Jaw Sideways	
AU 31:	Jaw Clencher	
AU 32:	Lip Bite	
AU 33:	Cheek Blow	
AU 34:	Cheek Puff	
AU 35:	Cheek Suck	
AU 36:	Tongue Bulge	
AU 37:	Lip Wipe	
AU 38:	Nostril Dilator	Nasalis, Pars Alaris
AU 39:	Nostril Compressor	Nasalis, Pars Transversa and Depressor Septi Nasi
AU 41:	Lid Droop	Relaxation of Levator Palpebrae Superioris
AU 42:	Slit	Orbicularis Oculi
AU 43:	Eyes Closed	Relaxation of Levator Palpebrae Superioris
AU 44:	Squint	Oricularis Oculi, Pars Palpebralis
AU 45:	Blink	Relaxation of Levator Palpebrae and Contraction of Orbicularis Oculi, Pars Palpebralis
AU 46:	Wink	Orbicularis Oculi
AU 51:	Head Turn Left	

AU 52: Head Turn Right
AU 53: Head Up
AU 54: Head Down
AU 55: Head Tilt Left
AU 56: Head Tilt Right
AU 57: Head Forward
AU 58: Head Back
AU 61: Eyes Turn Left
AU 62: Eyes Turn Right
AU 63: Eyes Up
AU 64: Eyes Down
AU 65: Walleye
AU 66: Cross-eye

I claim:

1. A Video Circuit comprising:

an information transfer device enabled to allow at least one user to send a communication to at least one viewer;

said information transfer device comprising an imaging system, one or more distribution channels, and a presentation system;

said imaging system being enabled to acquire sensory information from said user and from the environment of said user; said sensory information being capable of being subjected to at least one enhancement by said imaging system; said enhancement being at least one of the following: a change, a replacement, an augmentation, a modification, a re-texturing, a cleaning up of said sensory information, a deletion, a filtering, an override, a reposition, a re-staging;

said distribution channels being enabled to allow a communication from said user to said viewer;

said sensory information representing at least one of the following: the appearance, sound, motion, and characteristics of said user; the appearance, sound, motion, and characteristics of said environment of said user;

said user controlling said enhancement of said sensory information from said user and said environment.

2. The Video Circuit of claim 1, further including one or more libraries of formatting information describing specific methods and appearances for changing, enhancing, replacing, augmenting, modifying, retexturing, cleaning up, deleting, filtering, overriding, repositioning, restaging, or changing in a combination of such enhancements the sensory appearance of such users and/or such users' environments, where such formatting information may include such forms as software "plug-ins" (external subroutines), 2D images, 3D images, solid models, morph spaces, cyberspace environments, avatar costumes, augmentation props, and voice fonts, among others, and where such formatting information is selected by a person or by a computer program, transferred into said presentation system, and used by said presentation system along with said essential information in creating said sensory appearances of the one or more users and the users' environments.

3. The Video Circuit of claim 1, wherein one of said distribution channel(s) uses one or more of the following technologies: (a) the Internet; (b) a Local Area Network or Wide Area Network; (c) the telephone network; (d) computer tape; (e) the cellular telephone network; (f) CD-ROMs or DVD disks; (g) CDRs; (h) an Internet telephone; (i) cable typically used for cable TV; (j) fiber-optic cable; (k) radio waves, including the television broadcasting spectrum; (l) Web pages or FTP files

4. An Image Information Representation Subsystem comprising:

a) a means for accepting digitized sensory images of the scene,

b) a means for abstracting the essential information describing the appearances of the user(s) and/or their environment(s) composing the scene from the digitized sensory images,

c) a means of representing this essential information,

d) a means for making the represented essential information available for use or distribution.

5. the Image Information Representation Subsystem of claim 4 wherein said Image Information Representation Subsystem uses sound images and contains: (a) means to abstract the essential information in the speech sounds of the user; (b) means to abstract a voice font that describes the voice characteristics of the user; (c) a means of representing this essential information

6. The Image Information Representation Subsystem of claim 4 wherein one or more Image Acquisition Device(s) and one or more Means for Digitizing Images enable to take the sensory images from the Image Acquisition Device(s) or Means for Acquiring Images and convert them into a digital format for use by the Image Information Representation Subsystem

7. The Image Information Representation Subsystem of claim 4, further comprising a means of making the essential information available to a distribution channel whereby the one or more users, when using the Video Sender, can send sensory appearance essential information to the distribution channel or allow the channel to take the information, and whereby the one or more users can participate in the sending portion of a Video Sender conversation

8. Image Information Representation Subsystem of claim 6, further comprising:

a) the Imaging System wherein the system further includes a Fantasy Video email-sending engine having in addition: (a) an outgoing-message recording system that records image representation information from said image information representation subsystem into an outgoing e-mail message; (b) optionally, an outgoing-message storage buffer system into which said outgoing-message recording system records; (c) optionally, a re-record and review playback system that plays back a presentation of a previously-recorded outgoing message composed of said image representation information contained in said outgoing-message storage buffer system for user viewing, prompts for sending or deletion, and re-records the outgoing Video message if it is found to be unsuitable by the user; (d) an optional outgoing-message sending system that sends the e-mail to a distribution channel.

9. A Presentation Construction Subsystem comprising:

a) a means for accepting essential information describing the scene,

b) a means for creating sensory images from the essential information.

c) a means for making the created sensory images available for use or presentation

10. The Presentation Construction Subsystem of claim 9, further comprising:

one or more Presentation Device(s) that said sensory images to a viewer.

- CLAIMS 2 -

11. The Presentation Construction Subsystem of claim 9, wherein the presentation derived from the essential information and optional formatting information is constructed using one or more of the following technologies:

(a) the essential information includes the positions and orientations of key parts of the users' bodies or the environments

(b) the essential information includes the size and shape of key parts

(c) the essential information includes joint angles, actuator parameters, and Costume Configuration Vectors for key joints, sets of joints, and configurations in the users' bodies or in the environments

(d) the essential information includes routine calls in a graphics language that get interpreted or executed to help derive the presentation

(e) the essential information includes codes for selecting display components from various sets, including such things as the identities or recommended identities of augmentations and replacements for key parts of the users' bodies or the environments, etc

(f) the essential information includes points in a morph space, and the Presentation Construction Subsystem computes a regular morph or a perspective morph between different views to help construct the presentation

(g) the essential information uses codes derived from the Facial Action Encoding System to help determine the presentation

(h) the essential information includes a 3D model

(i) the essential information includes a Camera Information packet that specifies the locations or characteristics of virtual cameras used in helping to construct the presentation

(j) the essential information includes a Lighting Information packet that specifies the locations or characteristics of virtual light sources used in helping to construct the presentation

(k) the essential information includes a Texture Information packet that specifies the locations or characteristics of textures used in helping to construct the presentation

(l) the essential information includes a Literal Texture Information packet that specifies portions of one of the original acquired images to be used in helping to construct the presentation

(m) the essential information includes combinations of the above, which are used in combination to help construct the presentation

12. The Presentation System of claim 10, wherein one of said **presentation device(s)** comprises one or more of the following devices: (a) a computer monitor; (b) a television; (c) a high-definition television; (d) a flat-panel display, such as is mounted on a wall; (e) a 3-D head-mounted display; (f) a

system comprising a 3-D movie or computer monitor display, using lenticular lens gratings or LCD light-shutter devices in a flat panel or in viewers' glasses; (g) a hologram-making device; (h) a building-sized display sign; (i) a billboard; (j) a printer, color printer, photo printer, hologram film printer, hologram foil stamper, or color separation negative printer; (k) a picture-phone, screen phone, or videophone, including desktop and pay-phone styles; (l) a TV set-top device connected to a TV set or monitor, including cable boxes and family game computer systems; (m) a fax machine; (n) a cellular TV, picture-phone or videophone; (o) a wrist-watch TV or portable TV; (p) a wrist-watch picture-phone or videophone; (q) a laser-driven or N.C. router-based sculpting device, yielding output in wax, plastic, wood, metal, ice, or steel; (r) an LCD, dye, or plasma screen; (s) direct-to-magazine printers; (t) a laser-based device that projects an image directly onto the viewer's fovea; (u) a headset or wearable computer or fabric computer; (v) a window display on a vehicle such as an automobile, truck, bus, plane, helicopter, boat, tank, motorcycle, crane, etc.; (w) a neural transmitter that creates sensations directly in a viewer's body; (x) a computer-based movie projector or projection TV; (y) a hand-held game device; (z) a palmtop, laptop, notebook, or personal assistant computer; (aa) a screen display built into a seat or wall for use in the home, on airlines, inside cars, or in other vehicles; (bb) a computer monitor used in an arcade game or home computer game; (cc) a screen or speaker integrated with an appliance such as a refrigerator, toaster, pantry, or home-control system; (dd) any present or future device supporting a means for displaying a sensory presentation

13. the Presentation Construction Subsystem of claim 9, wherein said Presentation Construction System uses essential information describing the speech of the user and formatting information having a voice font, and has

(a) means to change the voice information by one or more of the following enhancements: replacing the voice font, augmenting the sound information with new information, modifying or filtering the existing sound information into something new, cleaning up or deleting parts of the sound information, overriding portions of the information with something different, repositioning the user's image in space, restaging the focus of the microphones, or a combination of such techniques;

(b) a means to generate an internal sound image, using a voice font and said essential information

whereby the true voice and sound environment of the user may be changed, enhanced, replaced, augmented, modified, retextured, cleaned up, deleted, filtered, overridden, repositioned, restaged, or changed in a combination of such enhancements, so that the viewer views (hears) an enhanced voice and sound environment, and so that the bandwidth requirements are relatively small.

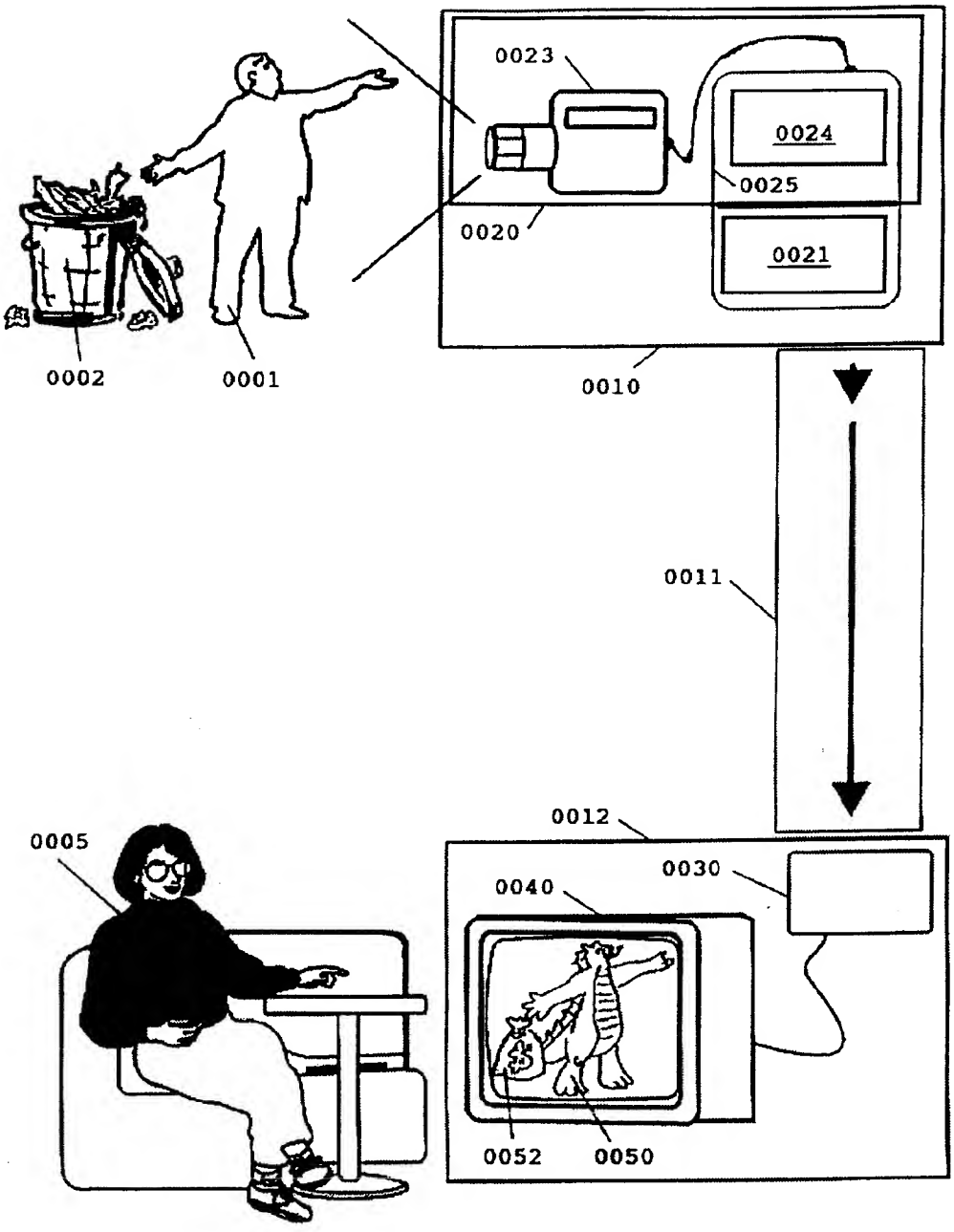

0023
0024
0025
0020
0021
0002
0001
0010
0011
0012
0005
0030
0040
0052
0050

Fig 1

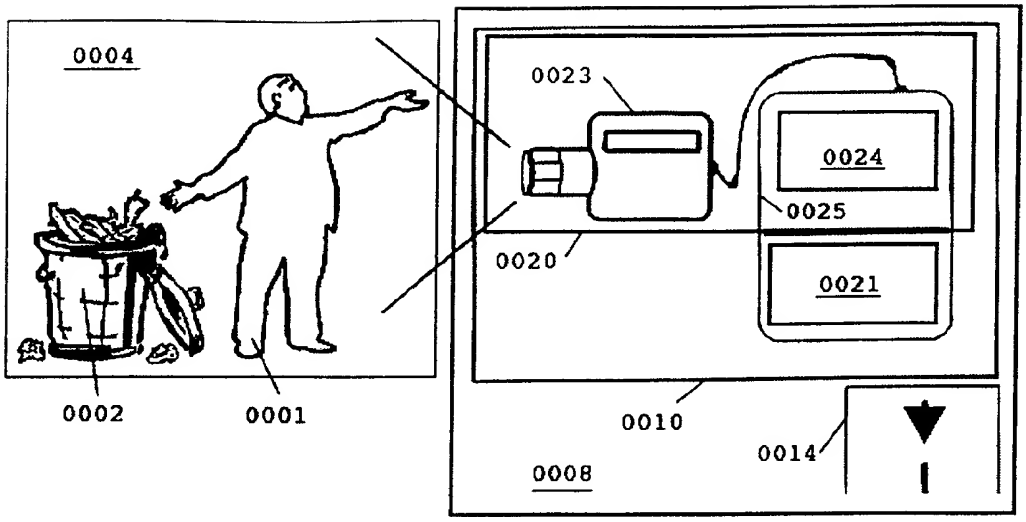

0004
0023
0024
0025
0020
0021
0002
0001
0010
0014
0008

Fig 1B

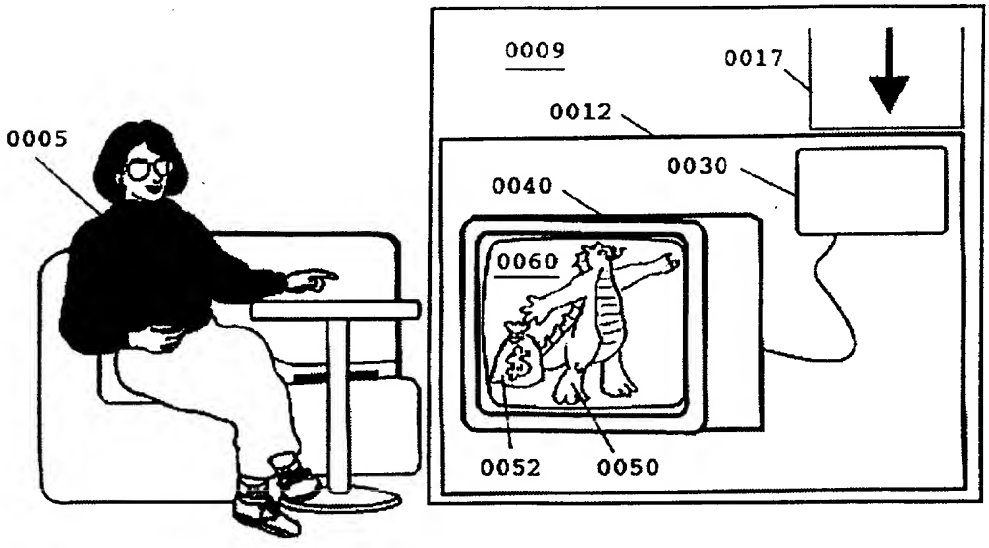

0009
0017
0012
0005
0030
0040
0060
0052
0050

Fig 1C

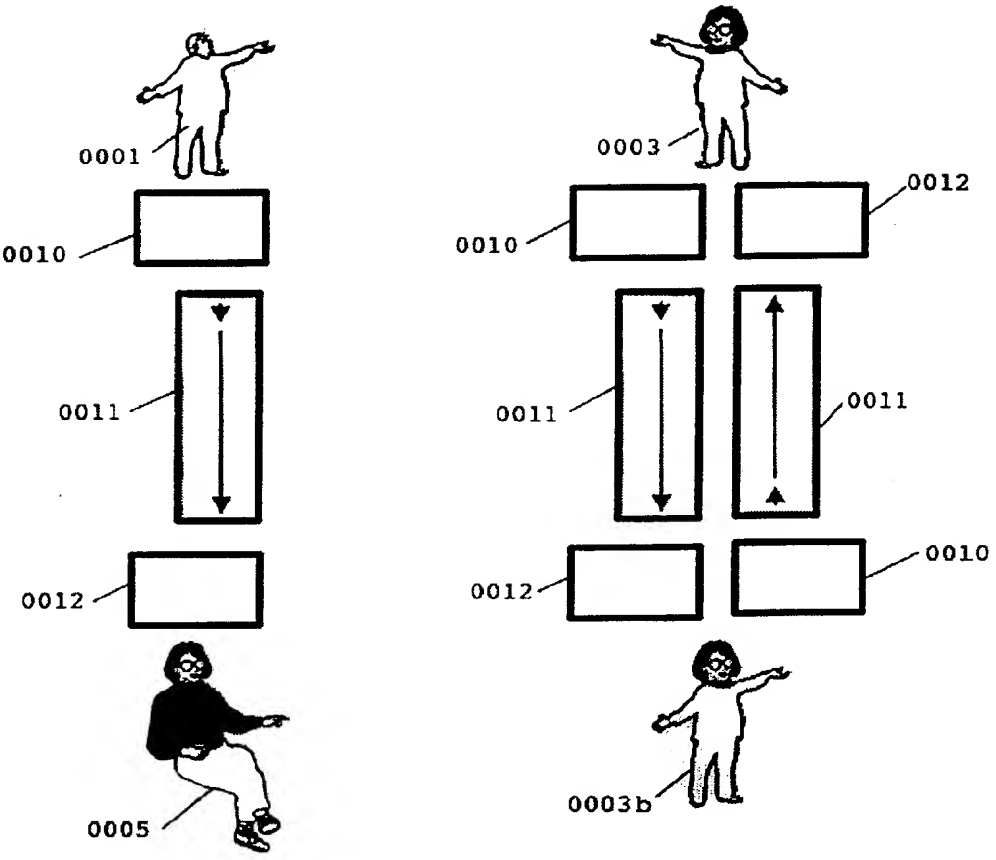

0001
0010
0011
0012
0005
0003
0012
0010
0011
0011
0012
0010
0003b

Fig 1D

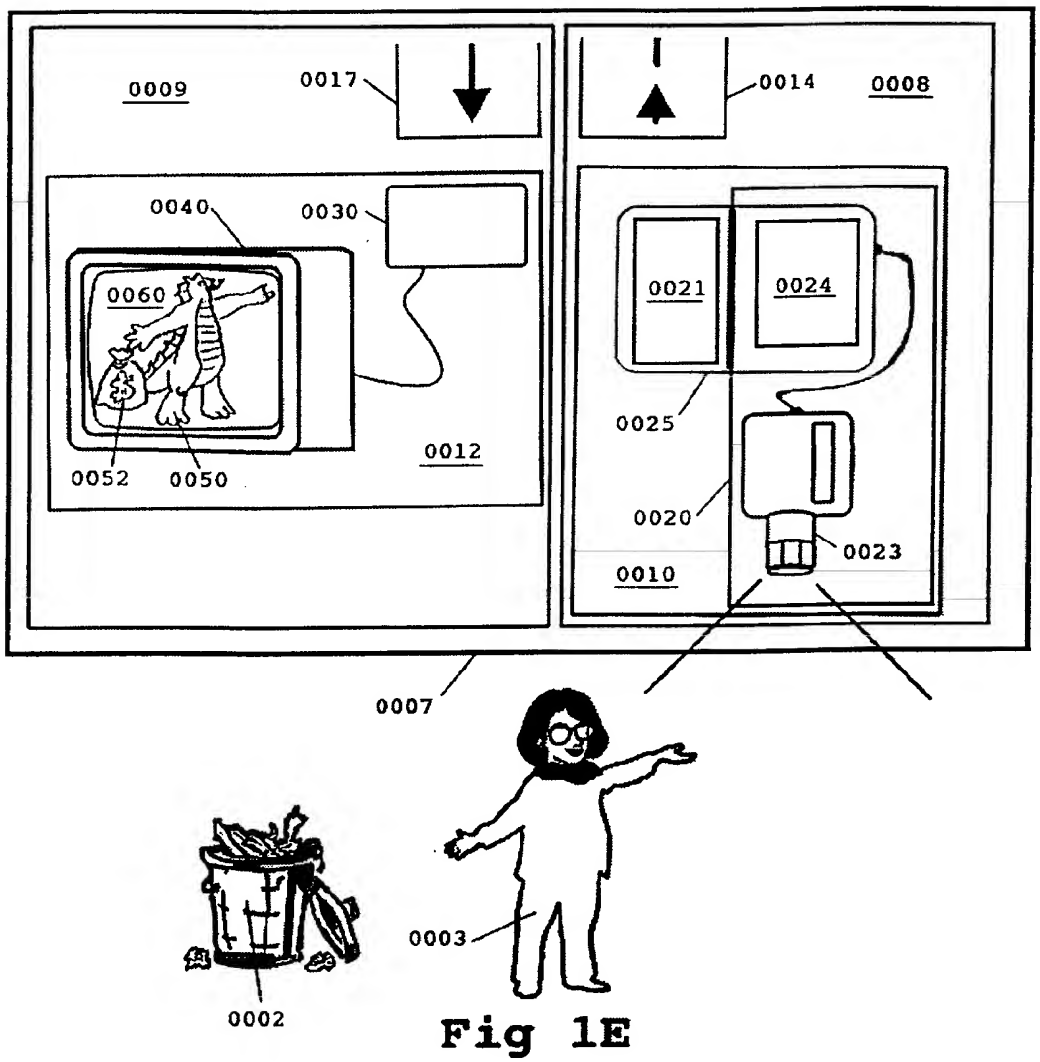

0009
0017
0014
0008
0040
0030
0060
0021
0024
0025
0012
0052
0050
0020
0023
0010
0007
0003
0002

Fig 1E

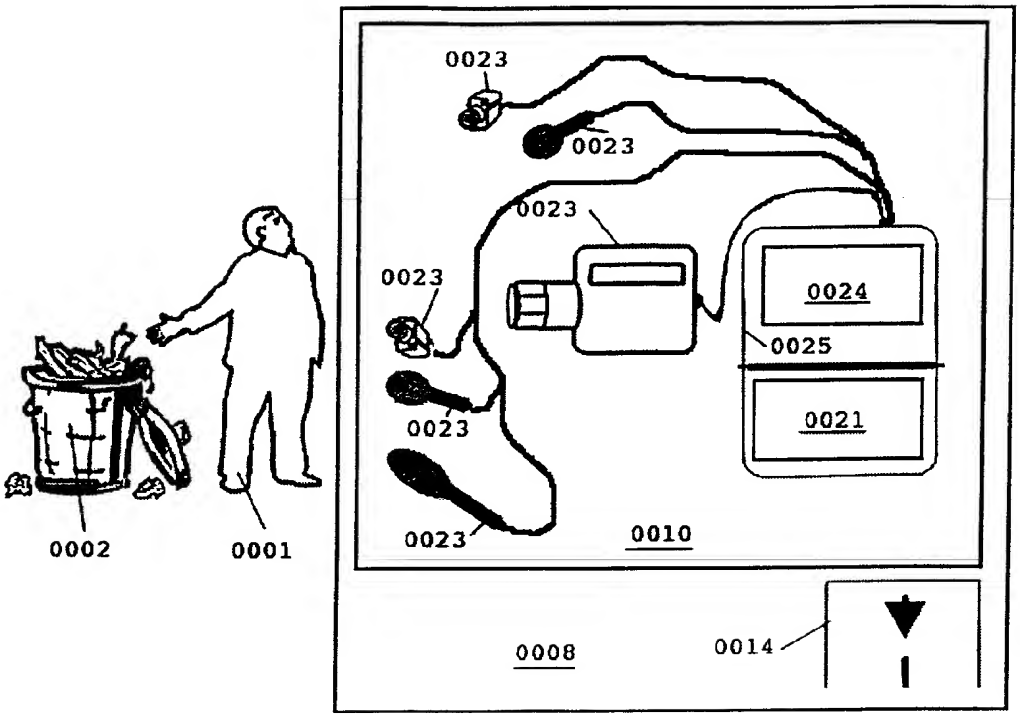

0023
0023
0023
0023
0024
0025
0021
0023
0023
0010
0002
0001
0008
0014

Figure 1F

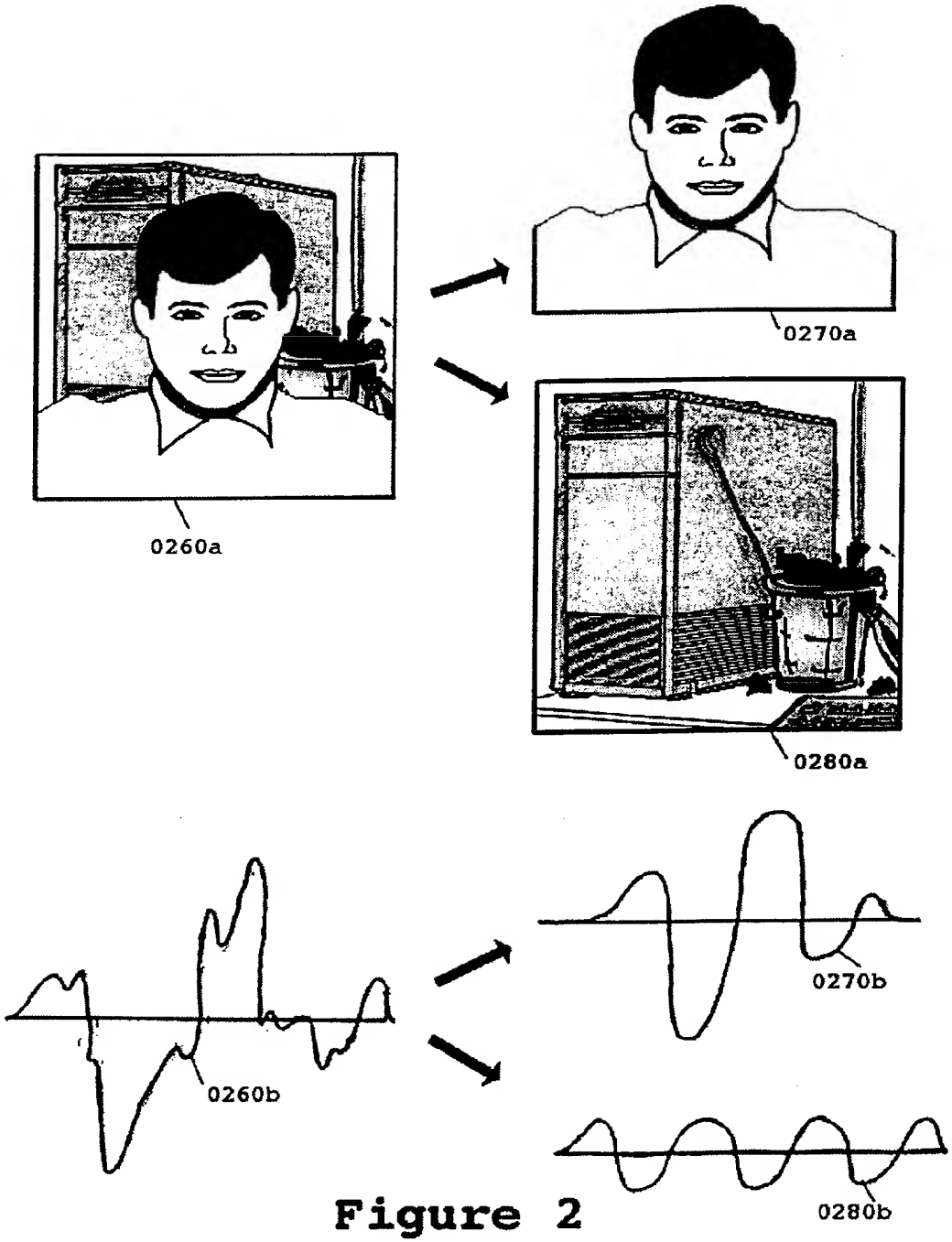


Figure 2

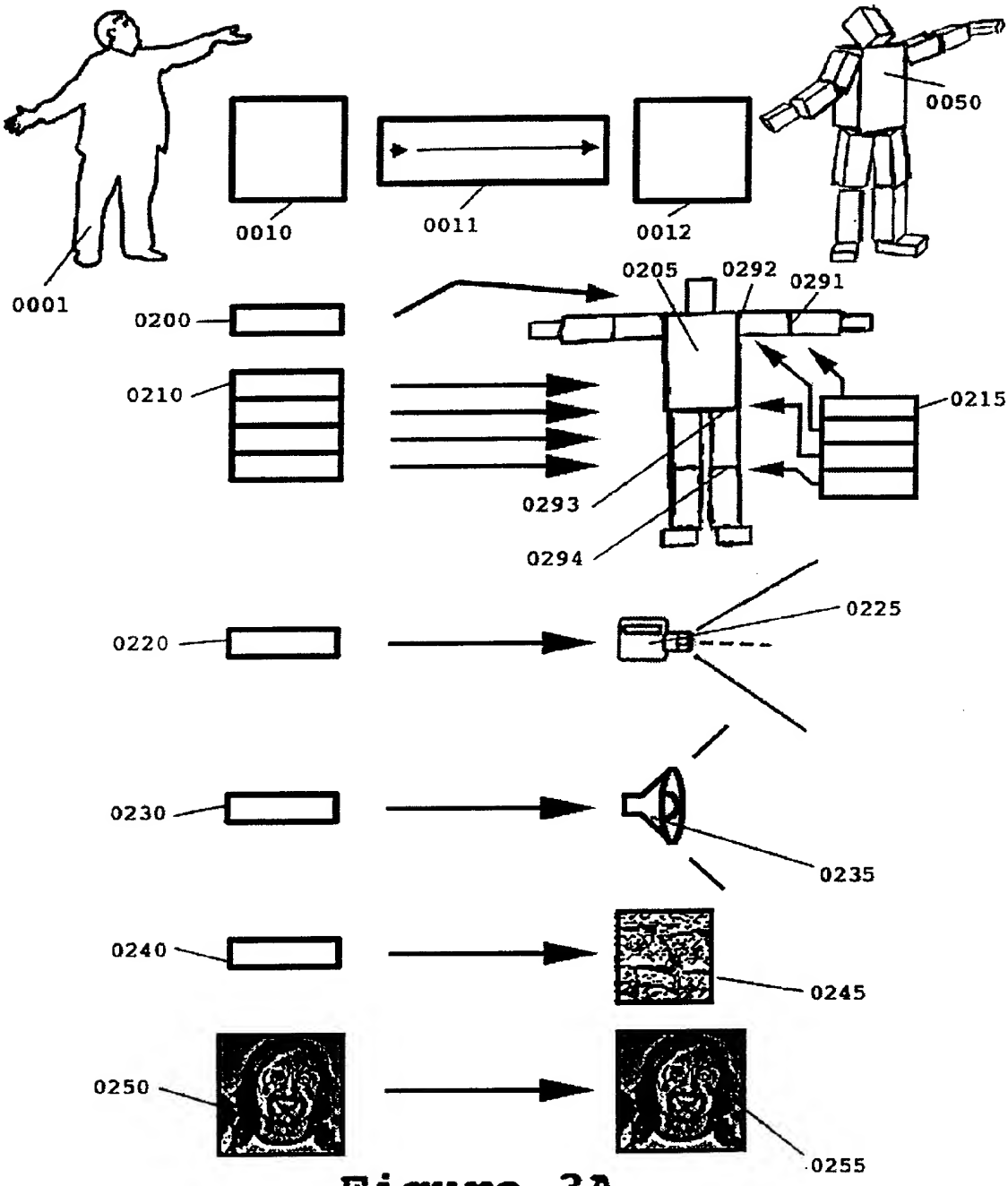

0001
0010
0011
0012
0050
0200
0205
0292
0291
0210
0215
0293
0294
0220
0225
0230
0235
0240
0245
0250
0255

Figure 3A

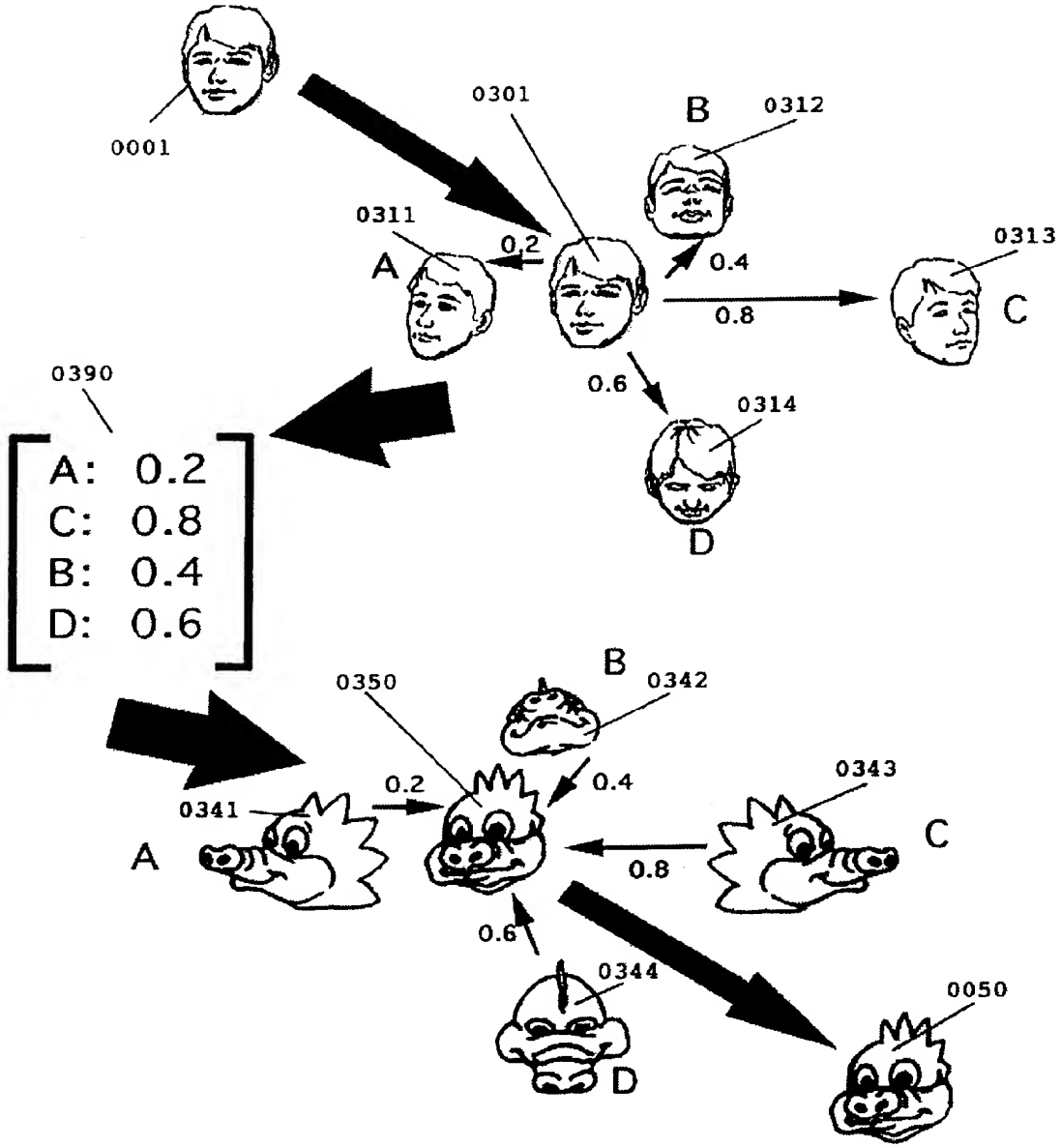

0001
0301
B
0312
0311
A
0.2
0.4
0313
C
0.8
0.6
0314
D
0390
A: 0.2
C: 0.8
B: 0.4
D: 0.6
0350
B
0342
0.2
0.4
0343
0341
A
C
0.8
0.6
0344
D
0050

Figure 3B

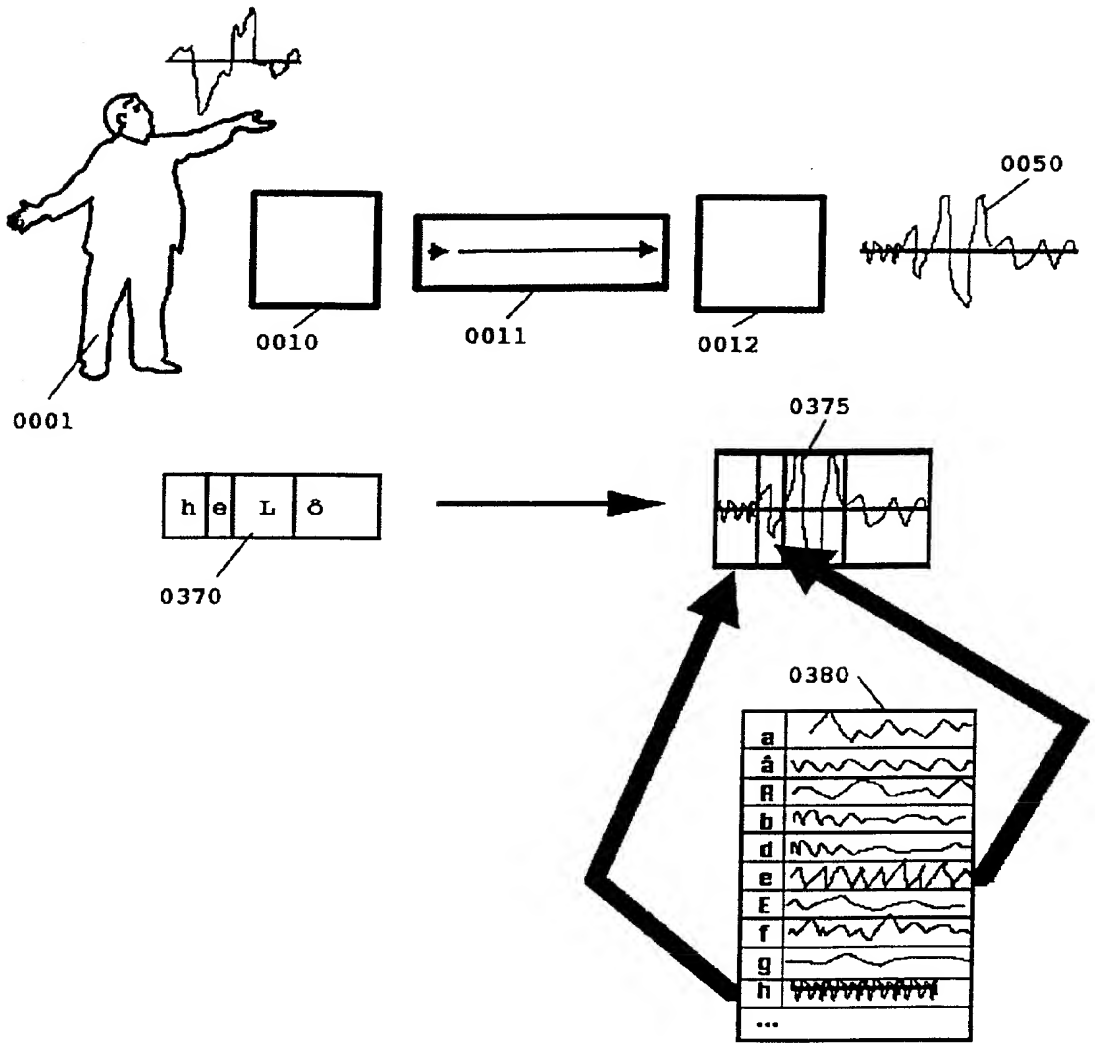


Figure 3C

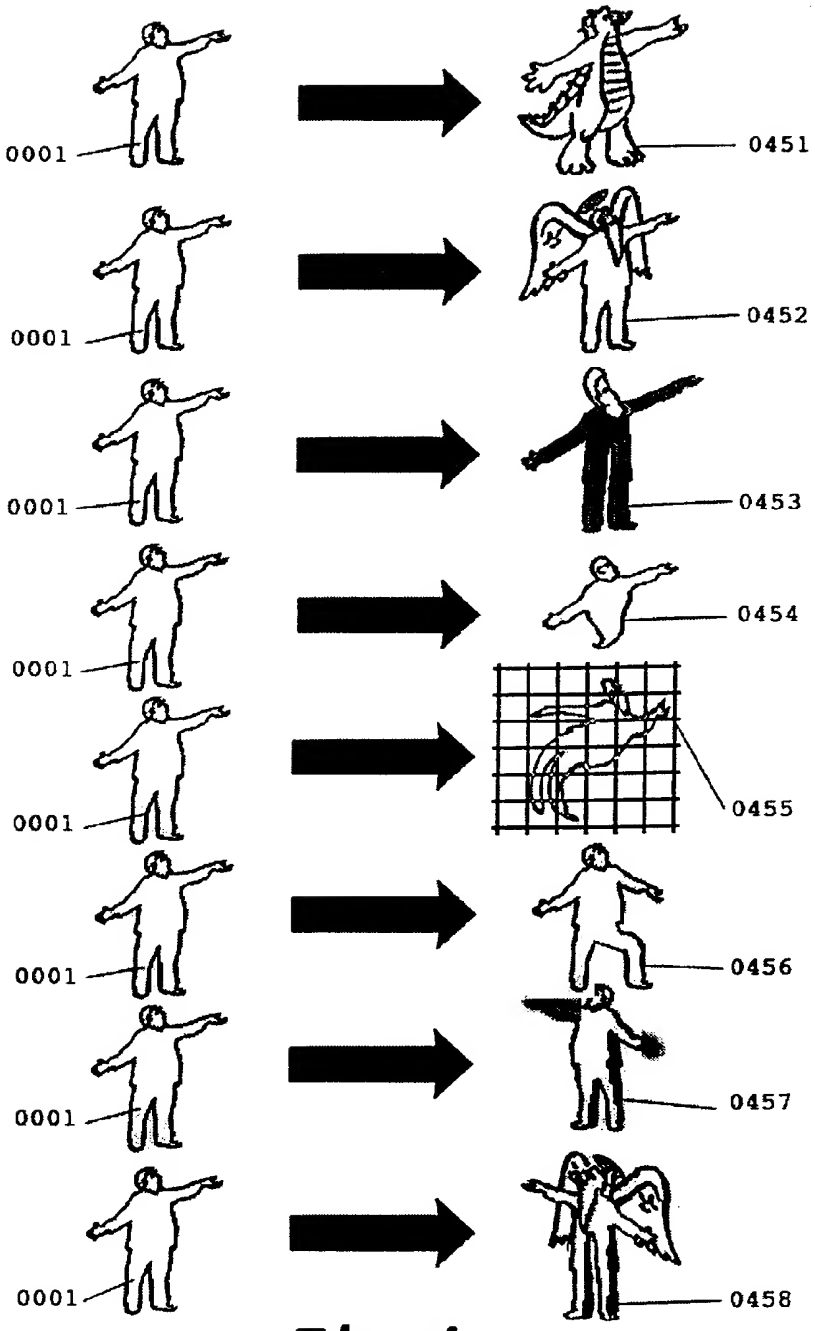

0001
0451
0001
0452
0001
0453
0454
0001
0455
0001
0456
0001
0457
0001
0458

Fig 4

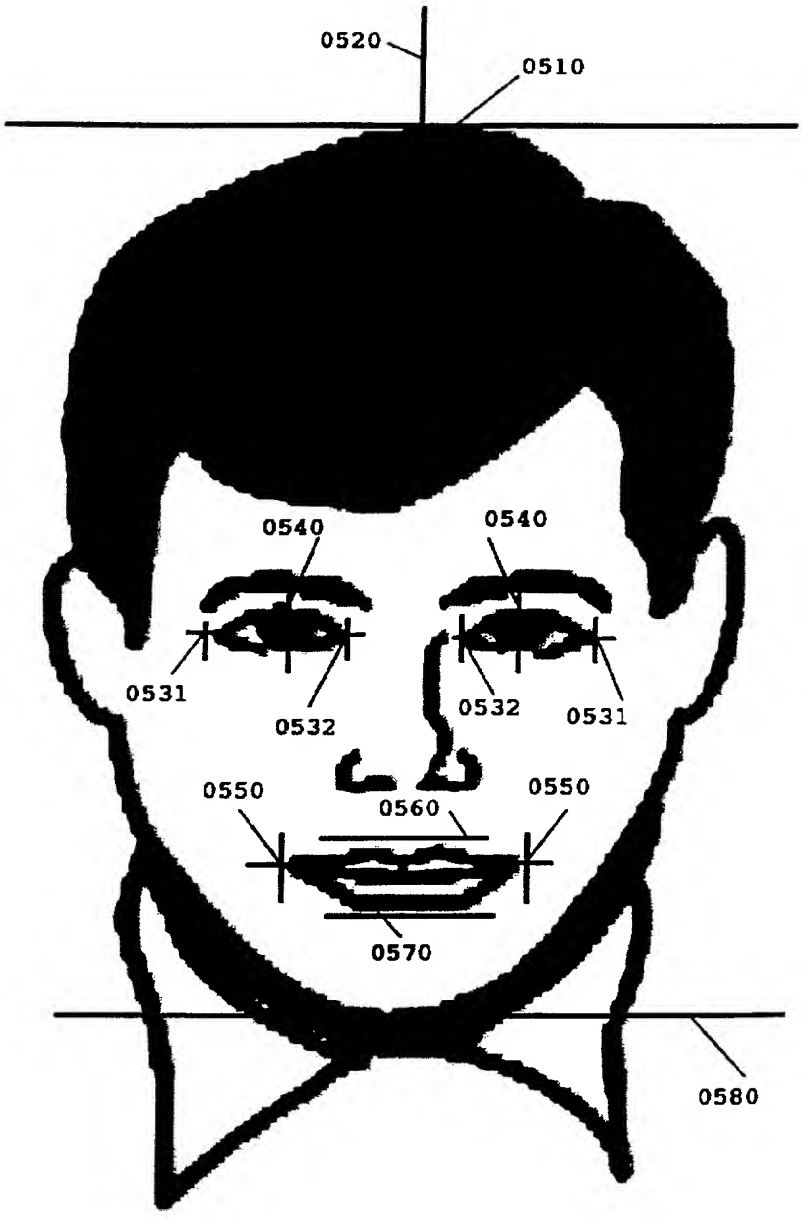

0520
0510
0540
0540
0531
0532
0532
0531
0550
0560
0550
0570
0580

Fig 5

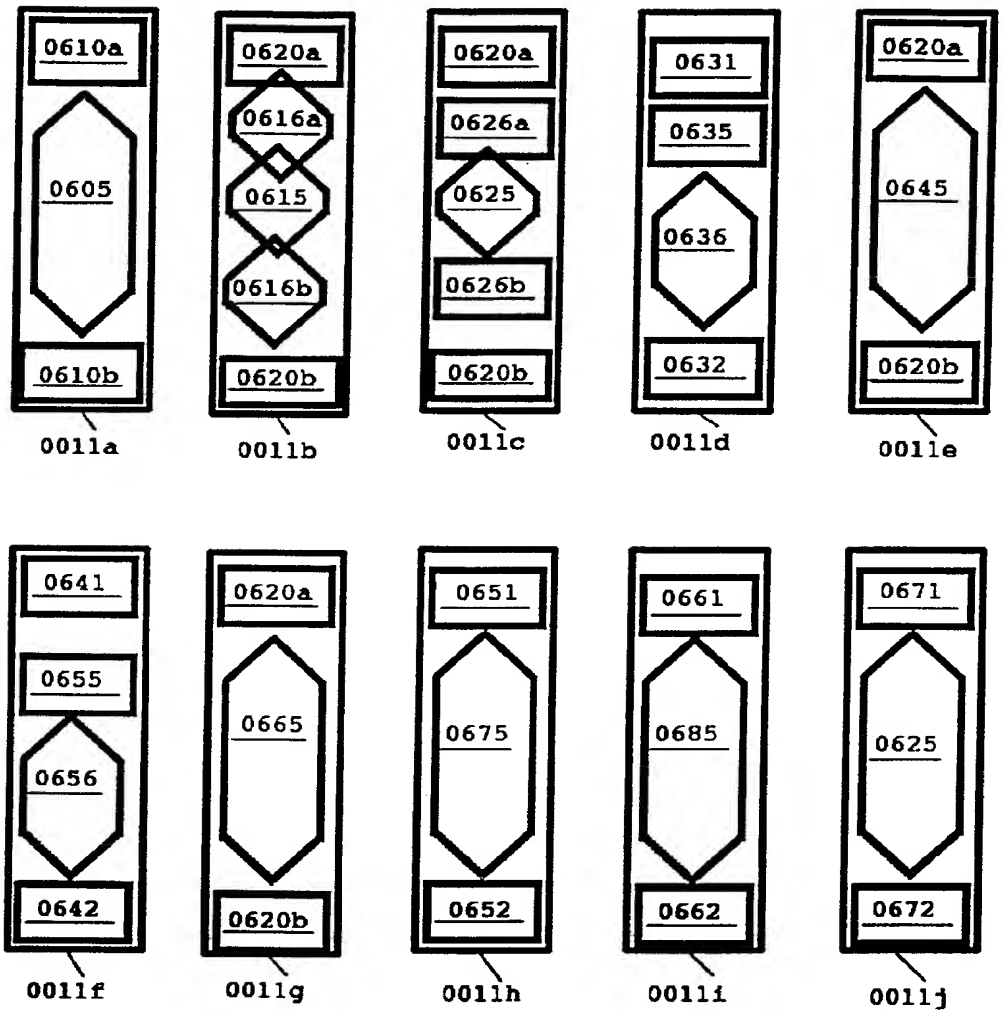

0610a
0605
0610b
0011a
0620a
0616a
0615
0616b
0620b
0011b
0620a
0626a
0625
0626b
0620b
0011c
0631
0635
0636
0632
0011d
0620a
0645
0620b
0011e
0641
0655
0656
0642
0011f
0620a
0665
0620b
0011g
0651
0675
0652
0011h
0661
0685
0662
0011i
0671
0625
0672
0011j

Fig. 6

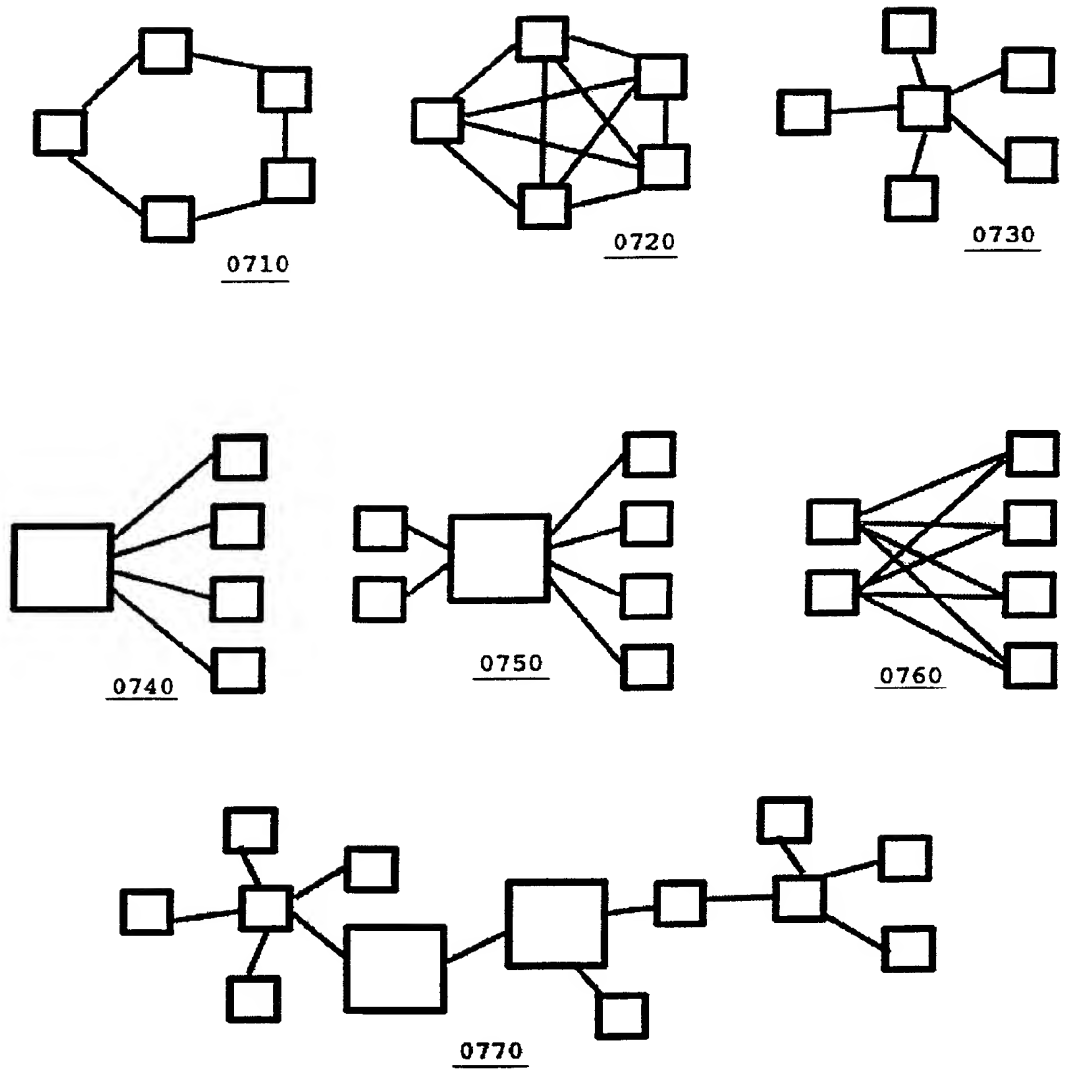


Fig 7

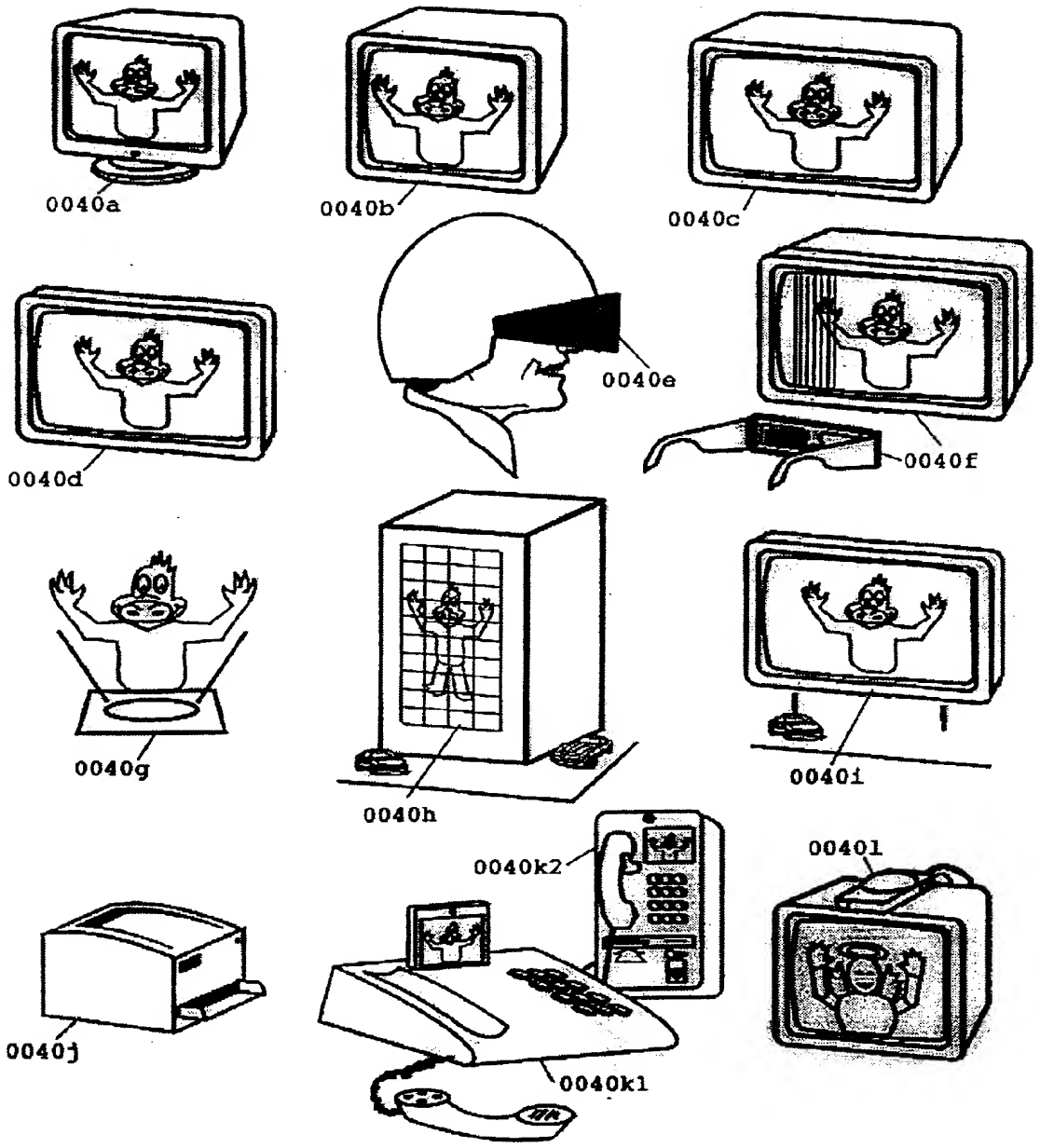

0040a
0040b
0040c
0040d
0040e
0040f
0040g
0040h
0040i
0040j
0040k1
0040k2
0040l

Fig 8A

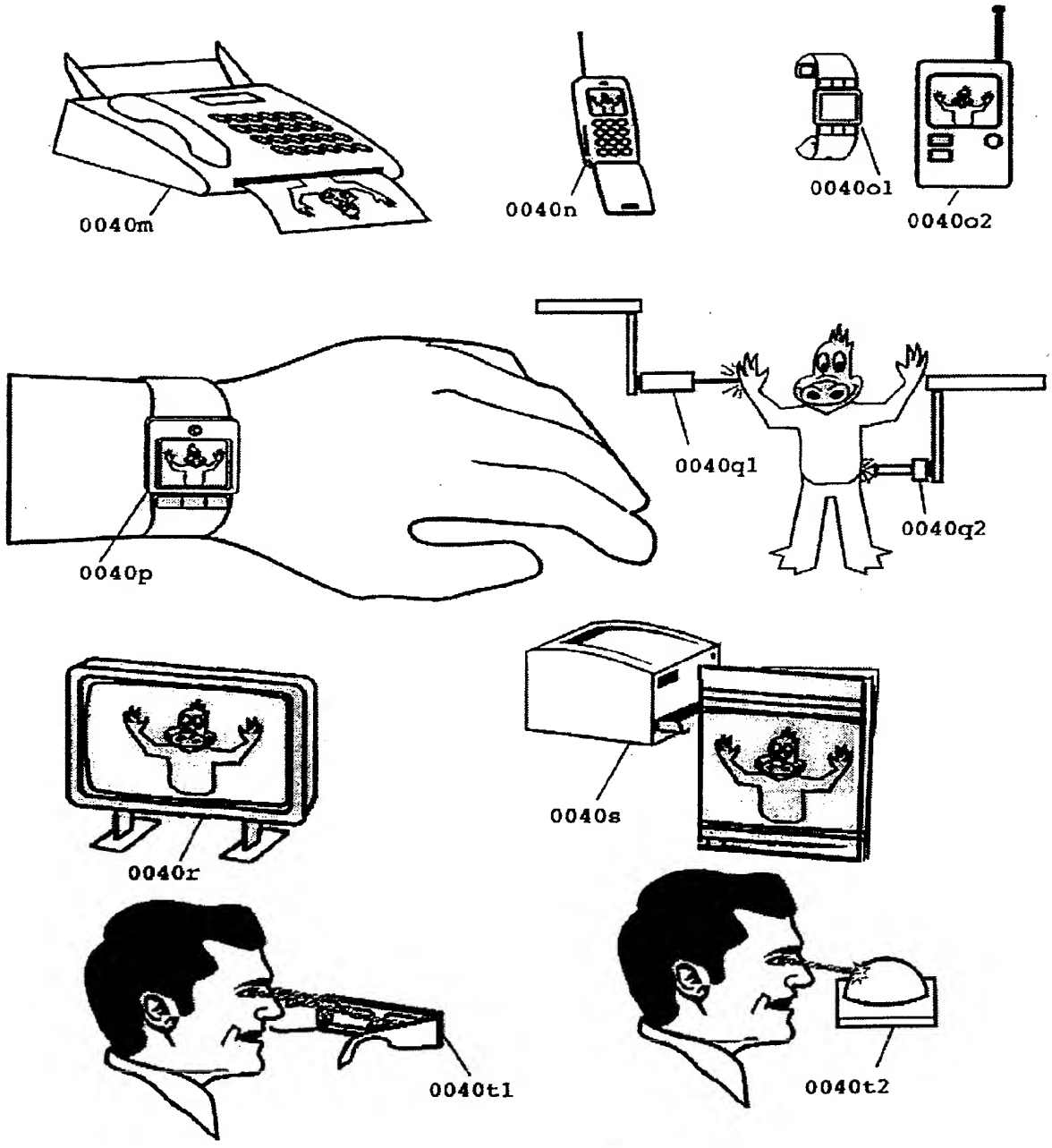

0040m
0040n
0040o1
0040o2
0040p
0040q1
0040q2
0040r
0040s
0040t1
0040t2

Fig 8B

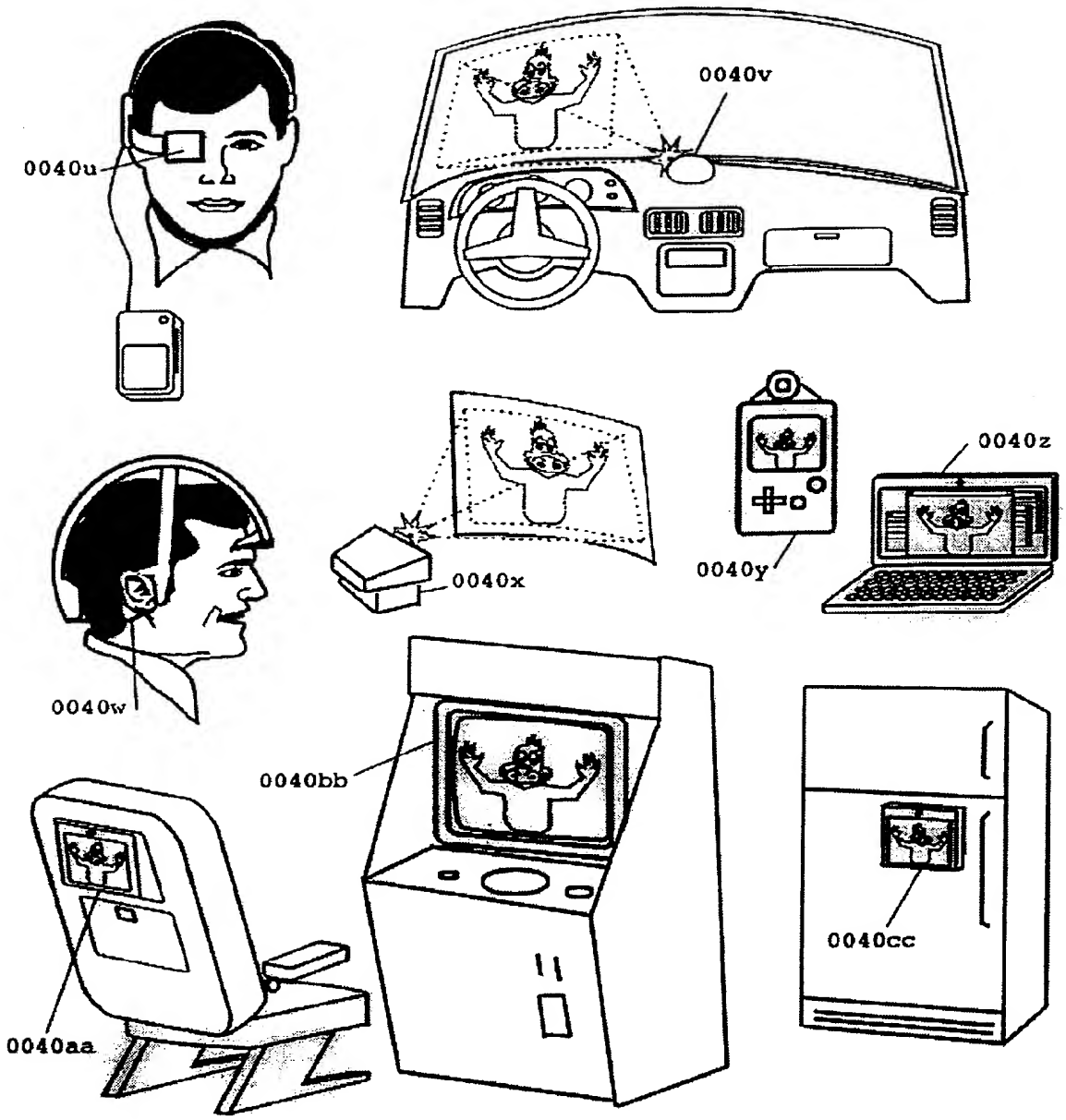


Fig 8C

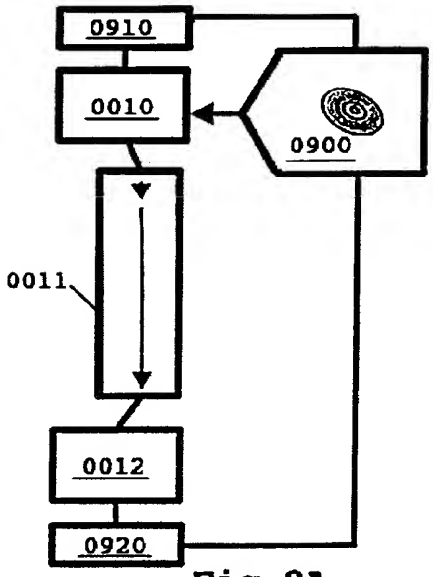


Fig 9A

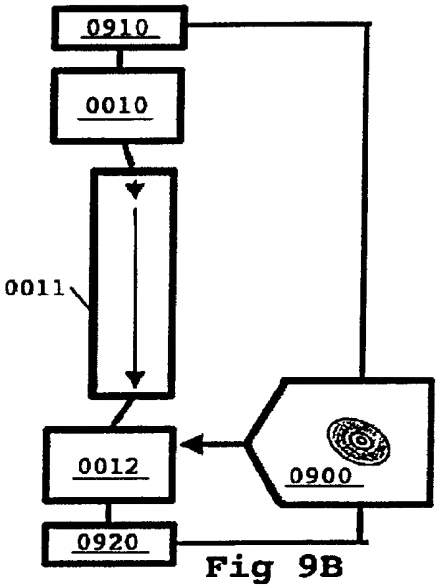


Fig 9B

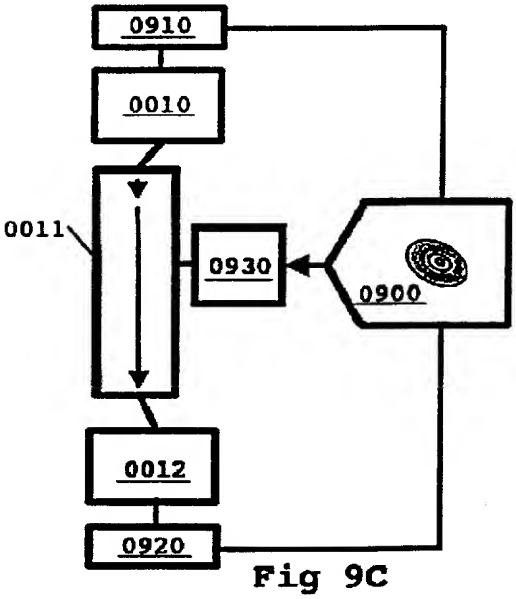


Fig 9C

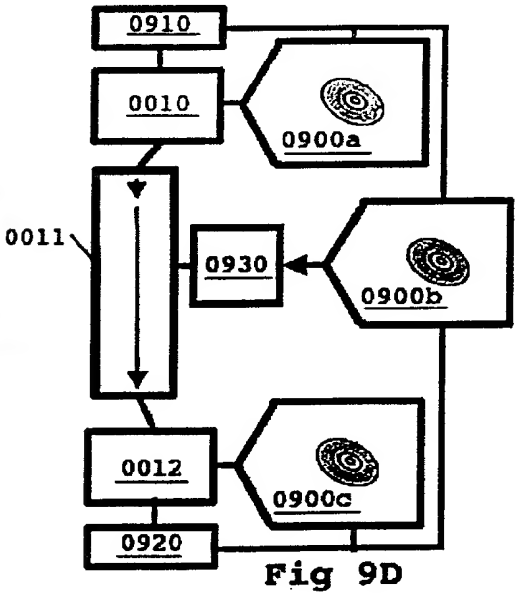


Fig 9D

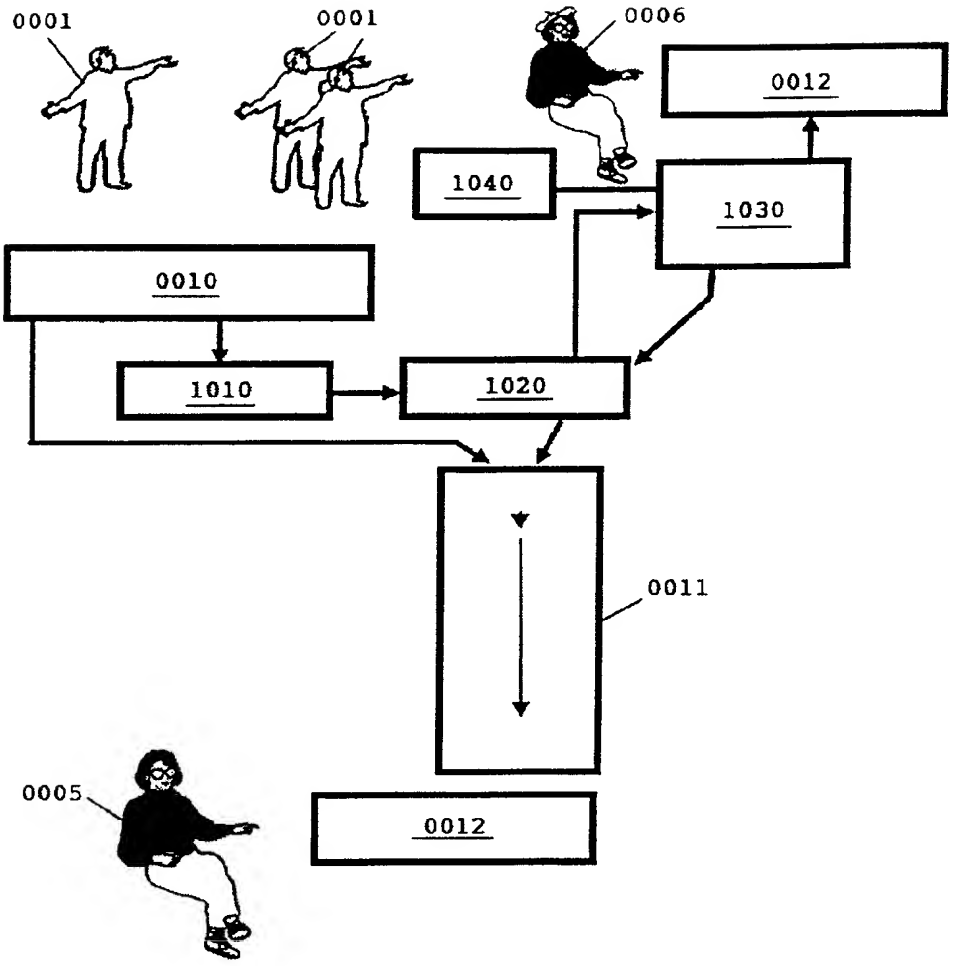


Fig 10

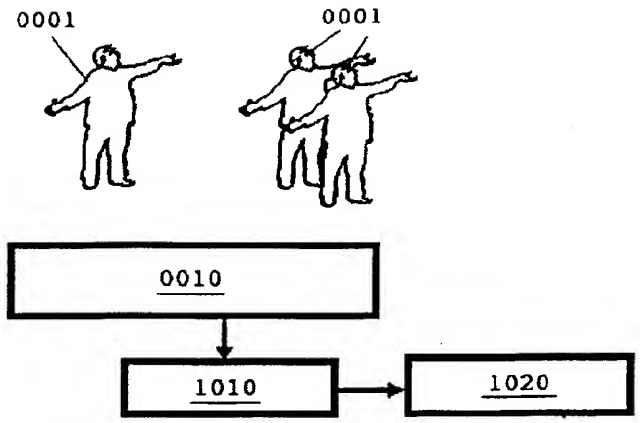


Fig 10B

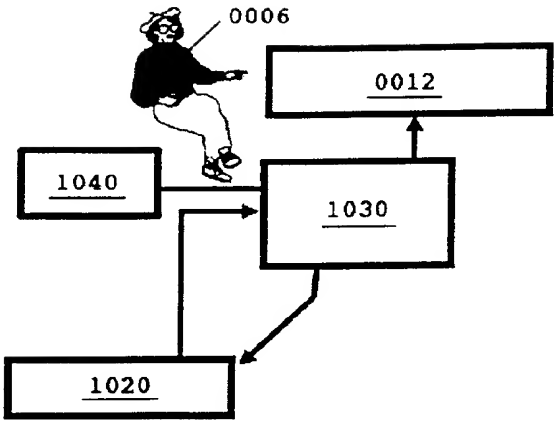


Fig 10C

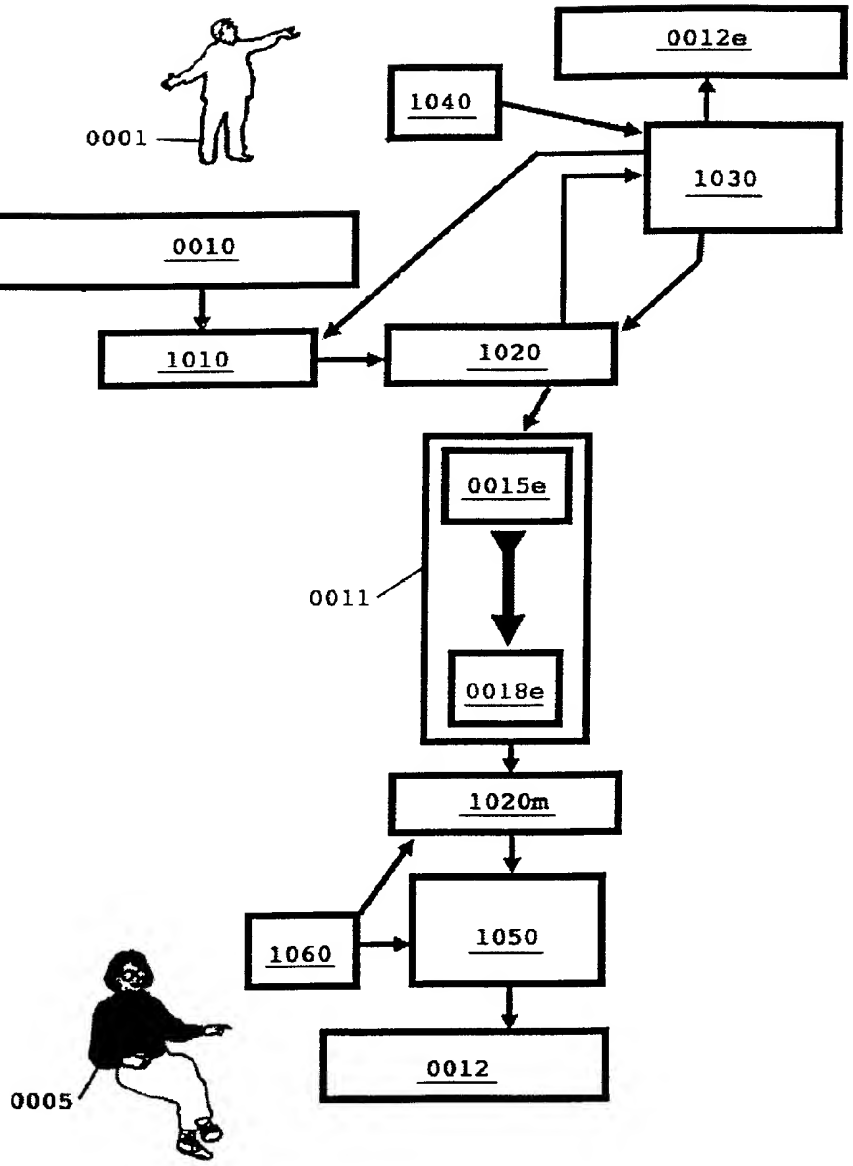

0001
0012e
1040
1030
0010
1010
1020
0015e
0011
0018e
1020m
1060
1050
0012
0005

Fig 11

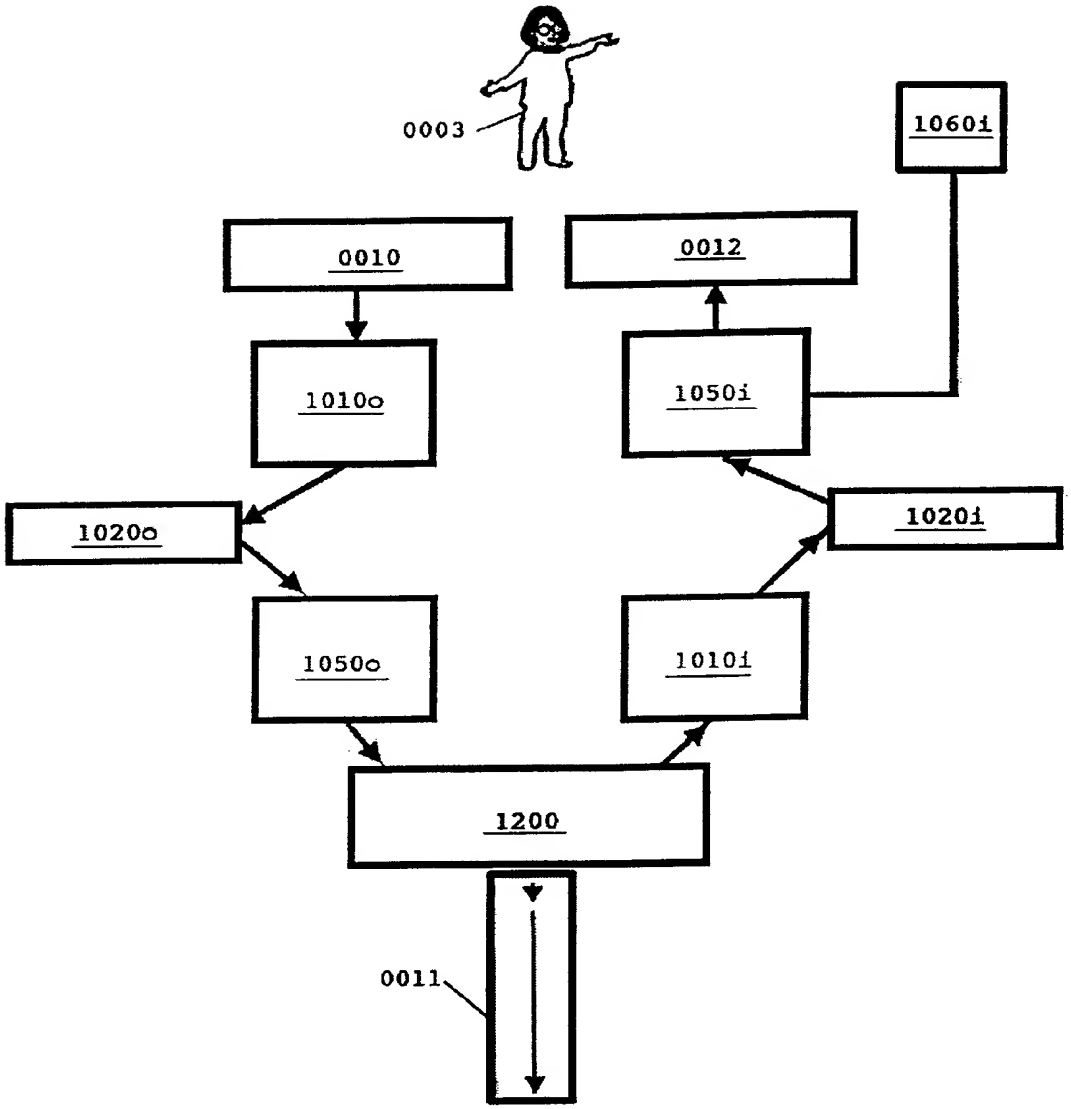

0003
1060i
0010
0012
1010o
1050i
1020o
1020i
1050o
1010i
1200
0011

Fig 12

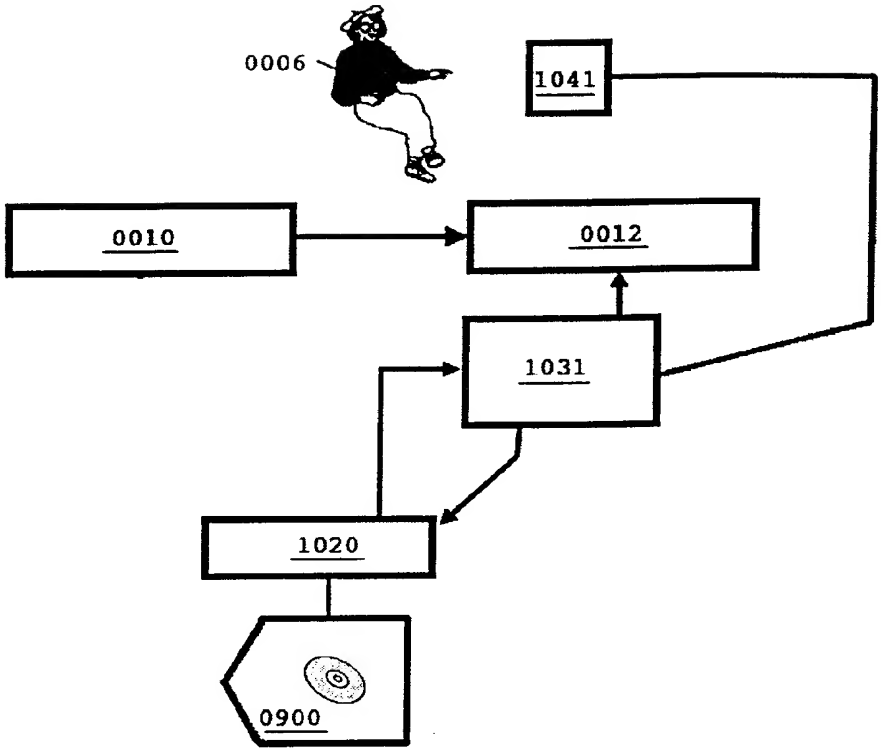


Fig 13

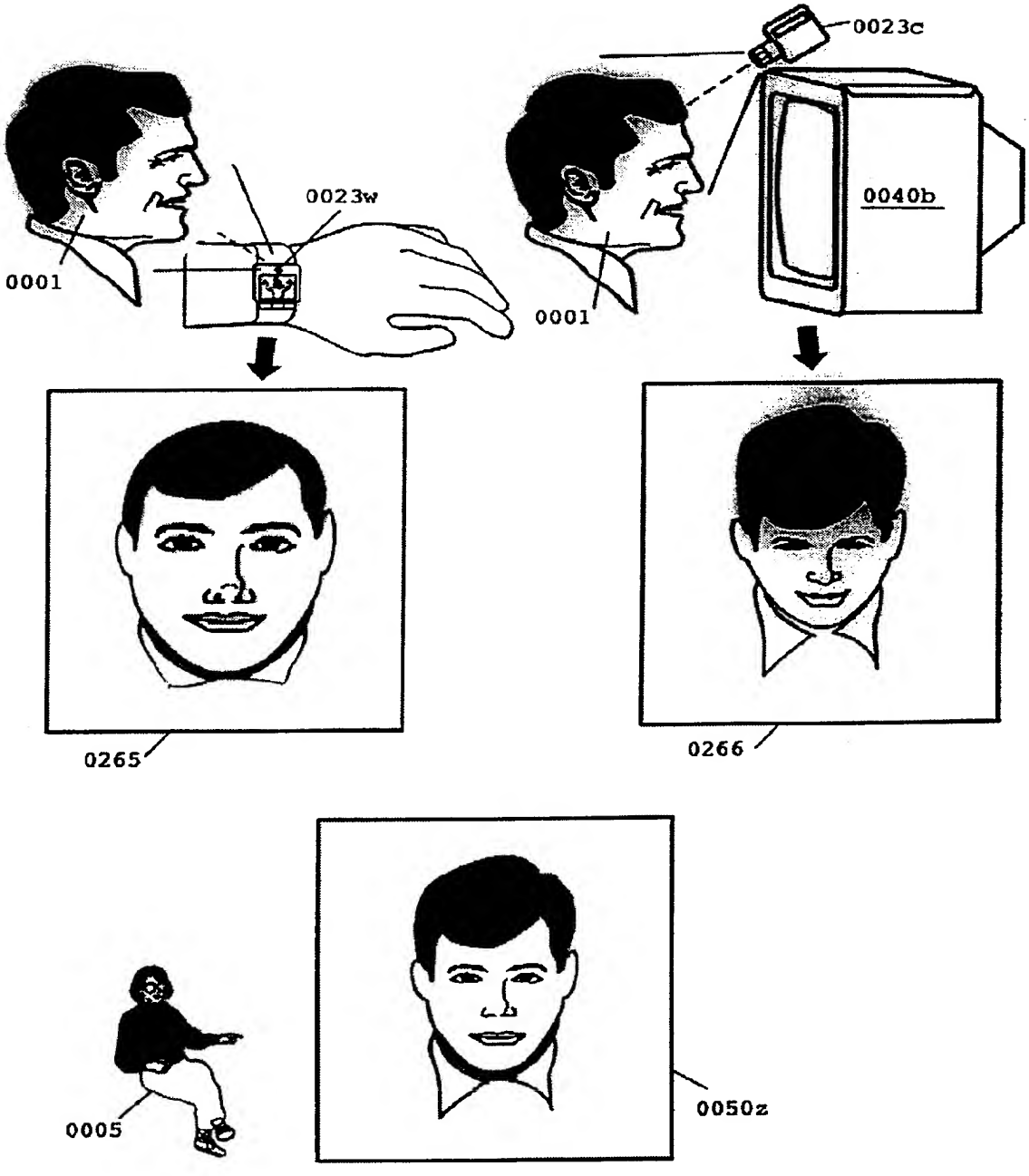


Fig 14

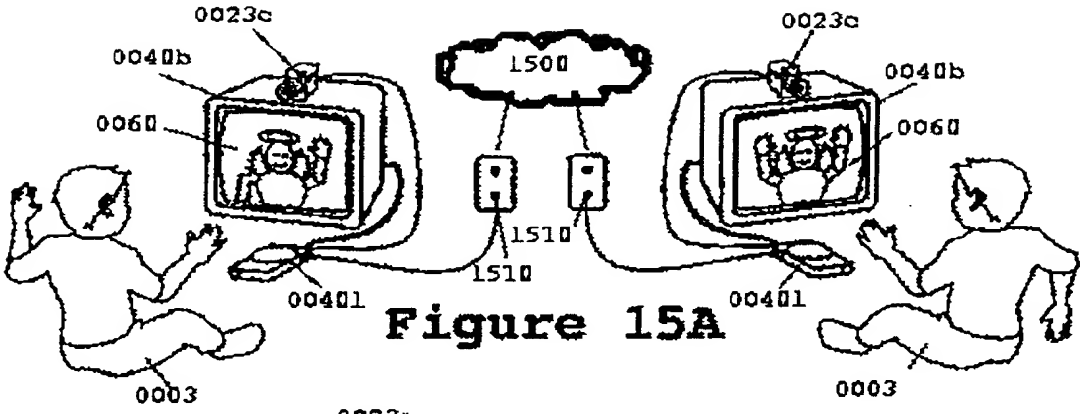


Figure 15A

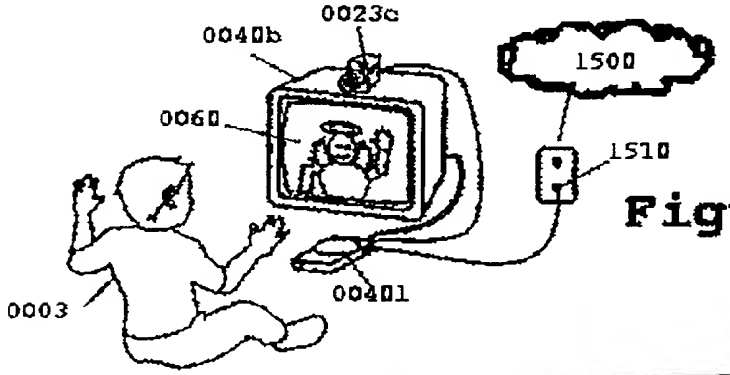


Figure 15B

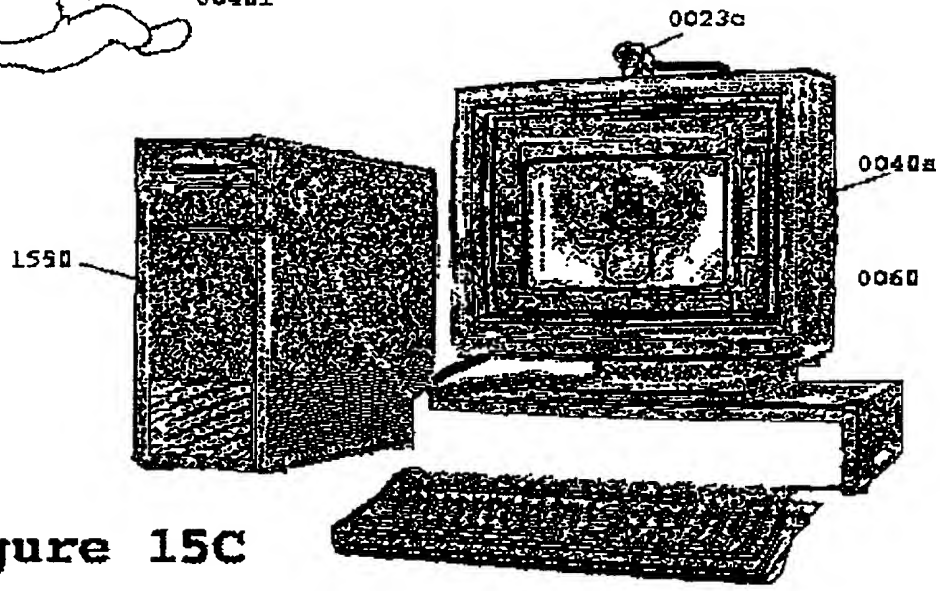


Figure 15C

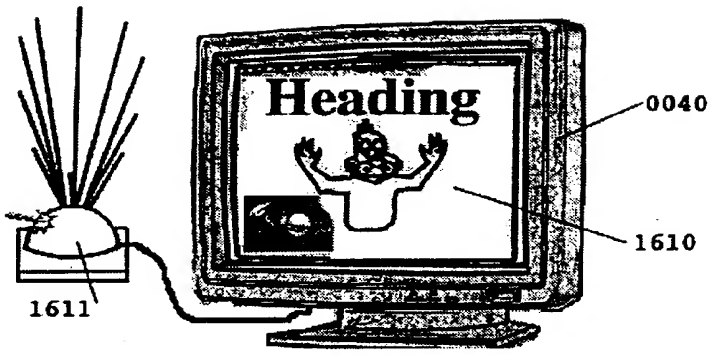


Figure 16A

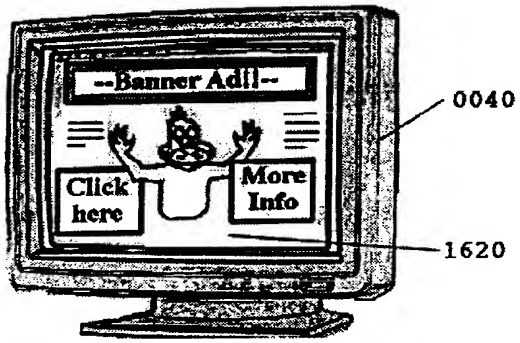


Figure 16B

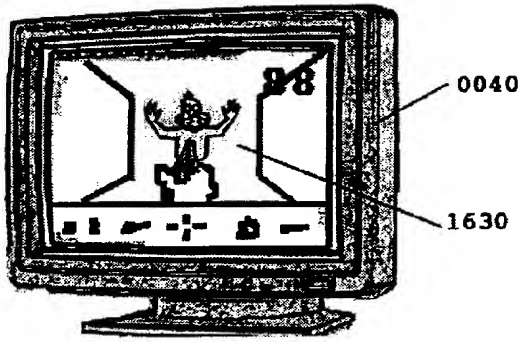


Figure 16C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/09515

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :H04N 7/14
US CL :348/14; 382/118
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 348/14,15,578,586,588,589; 382/118,209,232,243,254

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

None

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

None

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 5,596,362 A (ZHOU) 21 January 1997, see abstract, col.1, lines 60-67, col.5, lines 4-10, col.5, lines 66-67.	9-13 ---- 1-8
Y	US 5,659,625 A (MARQUARDT) 19 August 1997, abstract, col.6, lines 25-34,Fig.12G, col.64, lines 18-20.	1-7
Y	US 5,548,789 A (NAKANURA) 20 August 1996, col.3,lines 55-62, col.4,lines 9-20.	8
Y,P	US 5,896,128 A (BOYER) 20 April 1999, col. 20, lines 45-57, Fig 18, col. 24, lines 41-44.	1-7

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 20 JULY 1999	Date of mailing of the international search report 10 SEPT 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer PAUL LOOMIS Telephone No. (703) 305-4766

Form PCT/ISA/210 (second sheet)(July 1992)★